PCT INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/67, 15/70, 15/82, 15/29, 15/65, A01H 5/00, C07K 16/16

A1

(11) International Publication Number: **WO 98/21346**

(43) International Publication Date: 22 May 1998 (22.05.98)

(21) International Application Number: PCT/CA97/00799

(22) International Filing Date: 4 November 1997 (04.11.97)

(30) Priority Data:
08/747,788 14 November 1996 (14.11.96) US

(71) Applicant *(for all designated States except US)*: QUEEN'S UNIVERSITY AT KINGSTON [CA/CA]; Kingston, Ontario K7L 3N6 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KO, Kenton [CA/CA]; 1097 Lincoln Drive, Kingston, Ontario K7M 4Z5 (CA). PANG, Peng [CN/CA]; Apartment 911, 3440 Durocher Street, Montreal, Quebec H2X 2E2 (CA).

(74) Agents: STEEG, Carol, Miernicki et al.; Parteq Innovations, Queen's University, Kingston, Ontario K7L 3N6 (CA).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ENHANCED TRANSPORT WITH A PLASTID MEMBRANE TRANSPORT PROTEIN

(57) Abstract

A novel method to enhance translocation of molecules across or into cellular membranes using a plastid protein transport gene is described. The method can also be used to incorporate substances into membranes of organisms. Nucleic acid constructs include those which express a plastid protein transport protein or its equivalent in cells of all organisms.

BNSDOCID: <WO 9821346A1 >

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

EPIDOCID: WO 9921348A1 | >

ENHANCED TRANSPORT WITH A PLASTID MEMBRANE TRANSPORT PROTEIN

Background of the Invention

Plastids and mitochondria are double membrane-bound organelles found in eukaryotic cells. Chloroplasts, plastids containing the green pigment chlorophyll, are the most complex of the plant membranous organelles. Both chloroplasts and mitochondria specialize in the synthesis of ATP, using energy derived from electron transport from photosynthetic phosphorylation in chloroplasts and from oxidative phosphorylation in mitochondria.

To perform their role in the cell, plastids must continuously import all types of molecules, including proteins. The biogenesis and development of plastids require the coordinated assembly of plastidic- and nuclear-encoded proteins which are incorporated into membranes or other parts of the plastid. The process by which nuclear-encoded plastid proteins are targeted from the site of synthesis to the site of function is mediated by a complex series of events involving a multitude of proteinaceous signals and factors located in the cytosol and the plastidic compartment. Few of these factors are known and the molecular infrastructure underlying this important and complex event is far from being understood. Chloroplast envelope proteins play a major role in modulating the vectorial flow of molecules across the membrane, including large proteinaceous entities. The import of proteins into the plastid is a complex process requiring the close collaboration of both the outer envelope and the inner envelope membranes. Evidence for the possible existence of two distinct protein import complexes, one

in each envelope membrane, is beginning to emerge from a number of recent investigations (Waegemann, K. and Soll, J. (1991) *Plant J. 1*:149-158; Soll, J. and Waegemann, K. (1992) *Plant J. 2*:253-256; Schnell, D. and Blobel, G. (1992) *J. Cell. Biol. 120*:103-115; Alefson, H., Waegemann, K. and Soll, J. (1994) *J. Plant Physiol. 144*:339-345; Schnell, D., *et al.* (1994) *Science 266*:1007-1012; Kessler, F., *et al.* (1994) *Science 266*:1035-1039; Wu, C., Seibert, F.S. and Ko, K. (1994) *J. Biol. Chem. 269*:32264-32271). An important step in the characterization of the protein translocating complexes is the identification of the components involved. The identification of outer and inner plastid envelope polypeptides has been accomplished using a variety of strategies (Ma, Y., *et al.* (1996) *J. Cell Biol. 134*:315-327; Cornwall, K.L. and Keegstra, K. (1987) *Plant Physiol. 85*:780-785; Kaderbhai, M.A., *et al.* (1988) *FEBS Lett. 232*:313-316; Pain, D., *et al.* (1988) *Nature 331*:232-237; Schnell, D., *et al.* (1990) *J. Cell Biol. 111*:1825-1838; Hinz, G. and Flugge, U.-I. (1988) *Eur. J. Biochem. 175*:649-659; Soll, J. and Waegemann, K. (1992) *Plant J. 2*:253-256; Waegemann, K., *et al.* (1990) *FEBS Lett. 261*:89-92; Perry, S.E. and Keegstra, K. (1994) *Plant Cell 6*:93-105; Alefson, H., *et al.* (1994) *J. Plant Physiol. 144*:339-345; Schnell, D.J., *et al.* (1994) *Science 266*:1007-1012; Kessler, F., *et al.* (1994) *Science 266*:1035-1039; Wu, C., *et al.* (1994) *J. Biol. Chem. 269*:32264-32271; Hirsch, S., *et al.* (1994) *Science 266*:1989-1992; Seedorf, M., *et al.* (1995) *Plant J. 7*:401-411; Seedorf, M. and Soll, J. (1995) *FEBS Lett. 367*:19-22; Gray, J.C. and Row, P.E. (1995) *Trends Cell Biol. 5*:243-247). To date, these studies collectively indicate that envelope proteins with molecular masses of 21, 30, 34, 36, 44, 45, 51, 66, 70, 75 86, 97 and 100 kDa may be possible constituents of the plastid

BNSDOCID: <WO 9821346A1 I >

protein import apparatus; however, it is not obvious from the existing data whether some of the predicted similar sized components are identical to each other. Further, it is not known if any of the components have an active role in protein transport.

A mechanism for controlling the transport of substances into plastids could be used for modification of plastid pathways and products which occur in particular tissue types, such as the starch and fatty acid biosynthesis pathways in roots and seeds. Major drawbacks to plastid modification of this caliber, however, are the limited knowledge of genes encoding plastid transport proteins and the lack of characterization of such proteins.

Further, plastid transport mechanisms could be usefully incorporated into other organisms, especially prokaryotes. The heterologous production of protein pharmaceuticals in *Escherichia coli* is a cornerstone of the biotechnology industry. The technology provides an attractive and viable means for the production of proteins in quantities and qualities that are otherwise expensive and difficult to obtain from natural sources.

The gene sequence and encoded protein of one plastid membrane component has been identified. Ko, K., *et al.* (1995) *J. Biol. Chem.* 270:28601-28608; GenBank™/EMBL Data Bank, accession no. X79091. However, no role in transport was determined for this protein.

To date, no one has reported eukaryotic transport gene function or the functioning of a transport gene from a eukaryotic organelle in prokaryotic cells. An additional transport gene in both prokaryotic and eukaryotic cells would be useful to increase translocation and expression of cellular products. Increased

incorporation of proteins into membranes to elevate membrane function would also be desirable.

Summary of the Invention

This invention relates to a method for enhancing the transport of substances, particularly proteins, across a cellular membrane ("translocation") by means of isolated or recombinant nucleic acids encoding a plastid transport protein (Bce44B) or its functional equivalent. Nucleic acids which hybridize to the Bce44B gene are also encompassed by this invention when such hybridizing sequences encode the functional equivalent of the Bce44B protein. The present invention also relates to a method for enhancing the incorporation of substances, particularly proteins, into cellular membranes.

The cellular membranes can be those of prokaryotic or eukaryotic cells. They can include membranes of organelles, either single- or double-membrane bound organelles, as well as plasma membranes.

One object of this invention is to provide a method for the enhanced translocation and/or expression of the products of bacterial fermentation or culture. The nucleic acids described herein can be used to facilitate and increase the synthesis and secretion of products as a result of the enhancement of molecular transport in bacteria when encoded products of such nucleic acids are incorporated into the cellular membranes of bacteria.

Another object of this invention is to provide a membrane transport system which is independent of a naturally-occurring (native) transport system. Thus, prokaryotic or eukaryotic systems can be provided with selected transport mechanisms,

especially systems which bypass naturally-occurring transport mechanisms such as the SecA system in *E. coli*.

In another embodiment, the DNA of this invention can be used to enhance the growth of nonhuman organisms, and to produce useful quantities of many different substances. These substances include proteins and other molecules which are translocated by cells, as well as substances which are incorporated into cell membranes of all types: *i.e.*, plasma membranes, plastid membranes (including thylakoids), mitochondrial membranes (including cristae), Golgi membranes, endoplasmic reticula membranes, and the like.

Another object of this invention is to provide a vector comprising the DNA of SEQ ID NO:1 or a nucleic acid sequence which hybridizes to SEQ ID NO:1, and a promoter, which vector encodes membrane transport protein Bce44B, or a functional equivalent. Any hybridizing nucleic acid sequence capable of directing protein transport in a manner similar to Bce44B is included.

These vectors can be used in host cells such as prokaryotes and yeasts to enhance transport across cellular membranes. In addition, such vectors can be incorporated into the cells of nonhuman multicellular organisms to enhance translocation of substances across plasma membranes and/or organelle membranes.

Another object of this invention is the enhancement and modification of the import capability of the plastid compartment. Enhancement of protein import may increase the accumulation of all protein products in plastids, particularly in cells of seed and storage tissues. It can also boost the importation of enzymes involved in various biochemical pathways that function within the plastid. General enhancement of protein import in

plastids to elevate the amount of biochemical activity in plastids and the storage of proteins and possibly other valuable products can also be achieved using the methods of this invention. Enhancements of this nature can increase product synthesis.

The same benefits can be demonstrated in bacteria using this technology so that products based on bacterial secretion and exportation are more readily accumulated, solubilized and translocated. The chimeric genes and vectors exemplified for both bacteria and plants are available for introduction into selected strains. In fact, the methods of this invention are suitable for use in all prokaryotes and eukaryotes.

This invention also provides a method whereby the activity of the Bce44B protein or its functional equivalent can be or can produce a visual plant transformation marker. Thus, use of markers which are not plant in origin, or herbicide or antibiotic resistance markers which cause concerns about their regulation or widespread use in plants can be avoided.

Brief Description of the Figures

Figure 1 is a nucleic acid sequence which encodes the Bce44B protein (SEQ ID NO:1).

Figure 2 is the amino acid sequence of the Bce44B protein (SEQ ID NO:2) encoded by SEQ ID NO:1.

Figure 3 is a diagram of a Bce44B expression plasmid.

Figure 4 is a diagram of a truncated Bce44B expression plasmid pK117.

Figure 5 is a diagram of a truncated Bce44B expression plasmid pK118.

Figure 6 is a diagram of the construction of the CAMV-Bce44B-NOS transgene.

Detailed Description of the Invention

The invention relates to methods using isolated and/or recombinant nucleic acids (DNA or RNA) that are characterized by (1) their ability to hybridize to (a) a nucleic acid encoding a Bce44B protein or polypeptide, such as a nucleic acid having the sequence of SEQ ID NO:1 or (b) a portion of the foregoing (e.g., a portion comprising the minimum nucleotides required to encode a functional Bce44B protein; or by (2) their ability to encode a polypeptide having the amino acid sequence of Bce44B (e.g., SEQ ID NO:2), or to encode functional equivalents thereof; e.g., a polypeptide which when incorporated into a particular membrane (such as the outer envelope of a plastid) facilitates the transport of like molecules in the same manner as Bce44B; or by (3) both characteristics. A functional equivalent of Bce44B, therefore, would have a similar amino acid sequence and similar characteristics to, or perform in substantially the same way as, a Bce44B protein. A nucleic acid which hybridizes to a nucleic acid encoding a Bce44B polypeptide such as SEQ ID NO:1, can be double- or single-stranded. Hybridization to DNA such as DNA having the sequence SEQ ID NO:1, includes hybridization to the strand shown or its complementary strand.

In one embodiment, the percent amino acid sequence similarity between a Bce44B polypeptide such as SEQ ID NO:2 and functional equivalents thereof is at least about 25% ($\geq$ 25%). In a preferred embodiment, the percent amino acid sequence similarity between a Bce44B polypeptide and its functional equivalents is at least about 30% ($\geq$ 30%). More preferably, the

percent amino acid sequence similarity between a Bce44B polypeptide and its functional equivalents is at least about 40%, and still more preferably, at least about 45%.

Isolated and/or recombinant nucleic acids meeting these criteria comprise nucleic acids having sequences identical to sequences of naturally occurring Bce44B genes and portions thereof, or variants of the naturally occurring genes. Such variants include mutants differing by the addition, deletion or substitution of one or more nucleotides, modified nucleic acids in which one or more nucleotides are modified (e.g., DNA or RNA analogs), and mutants comprising one or more modified nucleotides.

Such nucleic acids, including DNA or RNA, can be detected and isolated by hybridization under high stringency conditions or moderate stringency conditions, for example, which are chosen so as to not permit the hybridization of nucleic acids having non-complementary sequences. "Stringency conditions" for hybridizations is a term of art which refers to the conditions of temperature and buffer concentration which permit hybridization of a particular nucleic acid to another nucleic acid in which the first nucleic acid may be perfectly complementary to the second, or the first and second may share some degree of complementarity which is less than perfect. For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions" and "moderate stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 (see particularly 2.10.8-11) and pages 6.3.1-6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, eds., Vol. 1, containing supplements up through Supplement 29, 1995), the

teachings of which are hereby incorporated by reference. The exact conditions which determine the stringency of hybridization depend not only on ionic strength, temperature and the concentration of destabilizing agents such as formamide, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, high or moderate stringency conditions can be determined empirically.

High stringency hybridization procedures can (1) employ low ionic strength and high temperature for washing, such as 0.015 M NaCl/ 0.0015 M sodium citrate, pH 7.0 (0.1x SSC) with 0.1% sodium dodecyl sulfate (SDS) at 50°C; (2) employ during hybridization 50% (vol/vol) formamide with 5x Denhardt's solution (0.1% weight/volume highly purified bovine serum albumin/ 0.1% wt/vol Ficoll/ 0.1% wt/vol polyvinylpyrrolidone), 50 mM sodium phosphate buffer at pH 6.5 and 5x SSC at 42°C; or (3) employ hybridization with 50% formamide, 5x SSC, 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2x SSC and 0.1% SDS.

By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson (1991) *Methods in Enzymology, 200*:546-556. Also, see especially page 2.10.11 in *Current Protocols in Molecular Biology* (*supra*), which describes how to determine washing conditions for

moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, from the lowest temperature at which only homologous hybridization occurs, a 1% mismatch between hybridizing nucleic acids results in a 1°C decrease in the melting temperature T_m, for any chosen SSC concentration. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined empirically for moderate or low stringency, depending on the level of mismatch sought.

Isolated and/or recombinant nucleic acids that are characterized by their ability to hybridize to (a) a nucleic acid encoding a Bce44B polypeptide, such as the nucleic acid depicted as SEQ ID NO:1, (b) the complement of SEQ ID NO:1, (c) or a portion of (a) or (b) (e.g. under high or moderate stringency conditions), may further encode a protein or polypeptide having at least one function characteristic of a Bce44B polypeptide, such as translocation activity (e.g., transport of β-lactamase across a bacterial cell membrane), or binding of antibodies that also bind to non-recombinant Bce44B. The catalytic or binding function of a protein or polypeptide encoded by the hybridizing nucleic acid may be detected by standard enzymatic assays for activity or binding (e.g., assays which measure the binding of a transit peptide or a precursor, or other components of the translocation machinery). Enzymatic assays, complementation tests, or other suitable methods can also be used in procedures for the identification and/or isolation of nucleic acids which encode a polypeptide such as a polypeptide of the amino acid sequence SEQ ID NO:2, or a functional equivalent of this

polypeptide. The antigenic properties of proteins or polypeptides encoded by hybridizing nucleic acids can be determined by immunological methods employing antibodies that bind to a Bce44B polypeptide such as immunoblot, immunoprecipitation and radioimmunoassay. PCR methodology, including RAGE (Rapid Amplification of Genomic DNA Ends), can also be used to screen for and detect the presence of nucleic acids which encode Bce44B-like proteins and polypeptides, and to assist in cloning such nucleic acids from genomic DNA. PCR methods for these purposes can be found in Innis, M.A., *et al.* (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA., incorporated herein by reference.

The nucleic acids described herein are used in the methods of the present invention for production of proteins or polypeptides which are incorporated into cellular membranes and which facilitate transport of substances across these membranes or incorporate these substances into the membrane. The "substances" can be proteinaceous molecules, such as proteins, peptides (including polypeptides), and molecules with peptide bonds, or can be nonpeptide compounds. In one embodiment, DNA containing all or part of the coding sequence for a Bce44B polypeptide, or DNA which hybridizes to DNA having the sequence SEQ ID NO:1, is incorporated into a vector for expression of the encoded polypeptide in suitable host cells. The encoded polypeptide consisting of Bce44B or its functional equivalent is capable of translocating substances, such as those described above. The term "vector" as used herein refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A vector, therefore, includes a plasmid or

viral DNA molecule into which another DNA molecule can be inserted without disruption of the ability of the molecule to replicate itself. The terms "translocating" or "translocation" mean the transport of substances across at least one cellular membrane from one part of the cell to another or into or out of the cell or organelle (*i.e.*, import or secret) or even into a periplasmic space (*i.e.*, export) as that found in *E. coli* between the inner and outer membranes.

Nucleic acids referred to herein as "isolated" are nucleic acids separated away from the nucleic acids of the genomic DNA or cellular RNA of their source of origin (e.g., as it exists in cells or in a mixture of nucleic acids such as a library), and may have undergone further processing. "Isolated" nucleic acids include nucleic acids obtained by methods described herein, similar methods or other suitable methods, including essentially pure nucleic acids, nucleic acids produced by chemical synthesis, by combinations of biological and chemical methods, and recombinant nucleic acids which are isolated. Nucleic acids referred to herein as "recombinant" are nucleic acids which have been produced by recombinant DNA methodology, including those nucleic acids that are generated by procedures which rely upon a method of artificial recombination, such as the polymerase chain reaction (PCR) and/or cloning into a vector using restriction enzymes. "Recombinant" nucleic acids are also those that result from recombination events that occur through the natural mechanisms of cells, but are selected for after the introduction to the cells of nucleic acids designed to allow or make probable a desired recombination event. Portions of the isolated nucleic acids which code for polypeptides having a certain function can

be identified and isolated by, for example, the method of Jasin, M., *et al.*, U.S. Patent No. 4,952,501.

A further embodiment of the invention is antisense nucleic acids or oligonucleotides which are complementary, in whole or in part, to a target molecule comprising a sense strand, and can hybridize with the target molecule. The target can be DNA, or its RNA counterpart (*i.e.*, wherein T residues of the DNA are U residues in the RNA counterpart). When introduced into a cell, antisense nucleic acids or oligonucleotides can inhibit the expression of the gene encoded by the sense strand or the mRNA transcribed from the sense strand. Antisense nucleic acids can be produced by standard techniques. See, for example, Shewmaker, *et al.*, U.S. Patent No. 5,107,065.

In a particular embodiment, an antisense nucleic acid or oligonucleotide is wholly or partially complementary to and can hybridize with a target nucleic acid (either DNA or RNA), wherein the target nucleic acid can hybridize to a nucleic acid having the sequence of the complement of the strand in SEQ ID NO:1. For example, an antisense nucleic acid or oligonucleotide can be complementary to a target nucleic acid having the sequence shown as the strand of the open reading frame of SEQ ID NO:1 or nucleic acid encoding a functional equivalent of Bce44B, or to a portion of these nucleic acids sufficient to allow hybridization. A portion, for example, a sequence of 16 nucleotides could be sufficient to inhibit expression of the protein. Or, an antisense nucleic acid or oligonucleotide complementary to 5' or 3' untranslated regions, or overlapping the translation initiation codon (5' untranslated and translated regions), of the Bce44B gene or a gene encoding a functional equivalent can also be effective. In another embodiment, the antisense nucleic acid

is wholly or partially complementary to and can hybridize with a target nucleic acid which encodes a Bce44B polypeptide.

In addition to the antisense nucleic acids of the invention, oligonucleotides can be constructed which will bind to duplex nucleic acid either in the gene or the DNA:RNA complex of transcription, to form a stable triple helix-containing or triplex nucleic acid to inhibit transcription and/or expression of a gene encoding Bce44B or its functional equivalent. Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) *Ann. Rev. Biochem.* *64*:65-95. Such oligonucleotides of the invention are constructed using the base-pairing rules of triple helix formation and the nucleotide sequence of the gene or mRNA for Bce44B. These oligonucleotides can block Bce44B-type activity in a number of ways, including prevention of transcription of the Bce44B gene or by binding to mRNA as it is transcribed by the gene.

Co-suppression refers to the overexpression of an endogenous or an introduced gene (transgene) wherein the extra copies of the gene result in the coordinate silencing of the endogenous gene as well as the transgene, thus reducing or eliminating expression of the trait. This technology can be used to limit transport of substances across membranes. See, for example, Jorgensen, *et al.*, U.S. Patents No. 5,034,323 and No. 5,283,184.

An alternative strategy to reduce the amount of Bce44B or its functional equivalent can be devised based on the use of "dominant-negative" mutant proteins. Certain types of mutations can be introduced into regulatory proteins that render them non-functional, but permit the mutant proteins to compete with the native proteins for their targets, which may include other protomers of a multimeric protein. Such competition by a non-functional protein means that overexpression of the mutant

protein can be used to suppress the activity of the native protein. In plants, dominant-negative strategies have been used successfully with other types of regulatory proteins. See, Boylan, M., *et al.* (1994) *Plant Cell 6*:449-460; Rieping, M., *et al.* (1994) *Plant Cell 6*:1087-1098; and Hemerly, A., *et al.* (1995) *EMBO J. 14*:3925-3936.

Those of skill in the art can appreciate the numerous examples wherein these strategies can result in commercial advantages in plants and other organisms. The following are several examples, not intended to be limiting, which describe the type of modification of plants that are possible using directed, tissue-specific expression. Seed or pollen development or production can be impaired or inhibited. This is especially useful where a company does not want seeds of a valuable crop, generated and harvested by competitors. The incorporation of truncated or mutant versions of Bce44B into organelle membranes could disturb transport to the extent that an antisense-type of inhibition occurs in a particular tissue, for example, resulting in male sterility. Expression in pollen which disrupts or impairs its function could be especially useful to prevent escape of transgenic materials from crops. Transfer of inhibitory genetic material to wild populations of unwanted plants could produce herbicidal effects and valuable weed control. This type of population control can extend to aquatic organisms, such as the dinoflagellates which produce red tides, or blue-green algae which are responsible for toxic blooms in freshwater systems. Even more useful, yields of crops harvested for vegetative materials could be increased by impairing the investment of energy in flower and seed production.

The invention also relates to methods using the proteins or polypeptides encoded by nucleic acids of the present invention. The proteins and polypeptides of the present invention can be isolated and/or recombinant. Proteins or polypeptides referred to herein as "isolated" are proteins or polypeptides purified to a state beyond that in which they exist in cells. "Isolated" proteins or polypeptides include proteins or polypeptides obtained by methods described herein, similar methods or other suitable methods, and include essentially pure proteins or polypeptides, proteins or polypeptides produced by chemical synthesis or by combinations of biological and chemical methods, and recombinant proteins or polypeptides which are isolated. Proteins or polypeptides referred to herein as "substantially purified" have been isolated and purified, such as by one or more steps usually including column chromatography, differential precipitation, or the like, to a state which is at least about 10% pure. Proteins or polypeptides referred to herein as "recombinant" are proteins or polypeptides produced by the expression of recombinant nucleic acids.

This invention provides a method for enhancing protein or peptide transport across a cellular membrane by incorporating an isolated nucleic acid encoding a plastid-derived transport protein, Bce44B (SEQ ID NO:1) or its functional equivalent into a cell or organism and maintaining the cell or organism under conditions appropriate for expression of the Bce44B protein or its functional equivalent.

The methods of the present invention can be used to enhance transport of molecules across any type of cellular membrane. This includes not only one or both of the double membranes of chloroplasts, mitochondria, nuclei and many bacteria, but also

ENSDOCID: <WO 9821346A1 | >

plasma membranes, thylakoids, cristae, vesicular membranes, golgi membranes, membranes which comprise the endoplasmic reticulum, and the like. In fact, artificially-made membranes or vesicles can also can also work when the appropriate proteins of the invention are incorporated, and be useful, for example, to sequester translocated compounds.

It is also understood that the transport proteins of the present invention can function to transport molecules across membranes of all types of cells, preferably plant cells, and in cell lines of organisms, preferably plants. The cells may comprise single-celled prokaryotic organisms; *i.e.*, the bacteria and cyanobacteria (blue-green algae). Alternatively, the cells can be animal-like, plant-like or fungal-like protists (single-celled eukaryotic organisms). Multicellular nonhuman organisms, including animals and all members of the fungi and plant kingdoms, are also suitable for application of the methods of this invention. Cell lines incorporating the enhanced transport methods of this invention can be derived from any of these eukaryotic organisms.

In addition to transport of molecules into and out of organelles and cells, the methods of the present invention can be used to incorporate proteins and other molecules into membranes. Like many proteins that form part of a membrane structure and facilitate transfer of molecules through the membrane, these proteins are also capable of integrating substances into membrane structure.

The invention described herein demonstrates for the first time that the Bce44B protein is an important membranous protein transport component which not only can translocate substances across membranes, but can incorporate molecules into membranes as

well. The transported substances can be proteins, molecules containing peptide bonds, or even nonproteinaceous molecules.

A novel aspect of this invention is the finding that this protein will incorporate and function in any membrane, including the inner and outer membranes of prokaryotes such as *E. coli*, not just the chloroplast envelope where it occurs naturally. Even more surprising is the discovery that the protein transports substances out of the bacterial cell; whereas, in a plastid, transport of proteins occurs from the cytosol into the intraorganellar sites of the organelle. That Bce44B is effective by itself, without requiring additional introduction of a multicomponent transport assembly, is both unusual and advantageous. Further, this is the first time a eukaryotic gene encoding a membranous protein transport component has been incorporated into a prokaryote and the protein expressed with functional activity.

In one embodiment, therefore, translocation of molecules is further enhanced by elevating the efficiency of molecular translocation and/or by creating additional pathways for transport via the introduction and expression of a central component of the plant plastid protein import apparatus in bacteria. The natural role of Bce44B in plants is to assist in the import of proteins into the plastid compartment. Expression of Bce44B in bacteria appears to stimulate a higher, more efficient level of protein translocation. While not to be limited by theory, these higher levels probably result from acceleration of the protein translocation process or from the addition of more pathways for transport. Further, the enhancement is probably stable because the newly introduced protein is not recognized as being bacterial in origin, thereby

reducing the possibility of down-regulation. This technology is novel because the enhancement of bacterial protein translocation is caused by a plant protein component and because the activity of this component also results in an elevation of protein expression. Together, these two features make a novel and significant impact on the ultimate levels of translocated proteins. The enhancement of foreign protein production and translocation can be achieved by stably introducing the desired gene construct into an *E. coli* strain expressing the plant protein component. The same method can also enhance transport into plastids and other organelles of plants at levels and/or in a manner which is not found in the naturally-occurring organism.

The chloroplast protein importation apparatus and Bce44B

The complex nature of protein translocation mechanisms observed in other membranous systems, such as the mitochondrion and the endoplasmic reticulum, suggests that there is most likely a significant number of plastid envelope components that need to be identified and characterized in detail. One major strategy for identifying and studying putative or possible components of the protein translocation apparatus is to isolate cDNA clones that encode all types of chloroplast envelope proteins and then to systematically sort out the identity and/or function of the clones. This approach allows skilled artisans to circumvent the technical problems and limitations of purifying small quantities of authentic proteins from the envelope.

To isolate the compounds for the methods of this invention, a cDNA clone encoding a 44 kDa envelope protein with unusual features was isolated and characterized. Ko *et al.* (1995) *J. Biol. Chem. 270*:28601. The 44 kDa polypeptide encoded by this

cDNA insert is a member of the Com44/Cim44 chloroplast envelope proteins (Wu *et al.* (1994) *J. Biol. Chem. 269*:32264). Although, these proteins have been found in close proximity to a partially translocated chimeric precursor protein, any role in the translocation process was heretofore unknown.

Specific antibodies were raised against the 44 kDa protein and used to determine the location of the immunologically-related polypeptides in the chloroplast envelope. The combined data from nucleotide sequencing, and RNA and protein blot analyses indicated the existence of multiple forms of the 44 kDa envelope protein. Depending on the plant species examined, immunologically-related protein bands with molecular masses of 42 to 46 kDa were observed. Organelle subfractionation, protease treatment and immunomicroscopy studies together provided an indication that the immunologically-related proteins could be present in both the outer and inner envelope membranes. Co-migration of the product synthesized from the cDNA insert with a 44 kDa immunoreactive band of the chloroplast envelope, and the *in vitro* import results, together suggest that the *in vitro* synthesized 44 kDa protein is targeted to the envelope membrane without any further processing.

Parallelism of chloroplast and bacterial protein translocation systems

Nuclear-encoded chloroplast precursor proteins are synthesized in the cytosol and then targeted to the organelle. The translocation of precursor proteins into the chloroplast is a highly complex process involving a multitude of components such as energy (mainly in the form of ATP), transit signals, proteinaceous envelope membrane factors, processing peptidases

and chaperones. These components are responsible for facilitating various steps of the import process which encompasses unfolding, specific binding to receptors on the outer envelope, translocation across the two envelope membranes, and precursor maturation (Keegstra, K. and Olsen, L.J. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol. 40*:471). A number of these chloroplastic components possess features in common with factors involved in the translocation of precursor proteins across the two bacterial cytoplasmic membranes and have indeed been demonstrated to be interchangeable to a certain extent. The transit signals of chloroplast precursor proteins can be recognized by the bacterial protein export machinery and transported into the periplasmic space, where processing to the correct mature molecular size occurs (Seidler, A. and Michel, H. (1990) *EMBO J. 9*:1743). In many cases, the processing sites of chloroplastic transit signals follow the bacterial -3, -1 rule and are readily cleaved by *E. coli* signal peptidase (Halpin *et al.* (1989) *EMBO J. 8*:3917). In addition to the prokaryotic nature of the thylakoid lumen-targeting signal, the 33 kDa subunit of the oxygen-evolving complex and plastocyanin traverse the thylakoid membrane via an azide-sensitive pathway resembling bacterial azide-sensitive protein export (Cline *et al.* (1993) *EMBO J. 12*:4105; Knott and Robinson (1994) *J. Biol. Chem. 269*:7843). Plastidic counterparts of the bacterial components GroEL, DnaK, SecA and SecY are present at internal sites in the organelle (Gutteridge and Gatenby (1995) *Plant Cell* 7:809; Nohara *et al.* (1995) *FEBS Lett. 364*:305; Reith and Munholland (1993) *Plant Cell* 5:465; Laidler *et al.* (1995) *J. Biol. Chem. 270*:17664) and in cases concerning GroEL, Srp54 and SecA counterparts, have been demonstrated to function in a very similar manner (Yuan *et*

BNSDOCID: <WO 9821346A1 I_>

al. (1994) *Science 266*:796; Gutteridge and Gatenby (1995) *Plant Cell* 7:809; Makai *et al.* (1994) *J. Biol. Chem. 269*:31338; Franklin and Hoffman (1993) *J. Biol. Chem. 268*:22175; Li, *et al.* (1995) *Proc. Natl. Acad. Sci. USA 92*:3789).

Expression of Bce44B in *E. coli*

The apparent parallelism of the chloroplastic and bacterial protein translocation systems suggests that other plastidic components are interchangeable and functional to some degree in a bacterial environment. It also suggests that translocation systems of a similar nature are present in membranes of other eukaryotic organelles and in cyanobacterial cells. Thus the methods of this invention provide a very useful and novel approach for studying different aspects of the interchanged membrane transport components. This idea was tested by introducing and expressing the recently identified component of the chloroplast protein import apparatus, Bce44B, in bacteria. Bce44B was previously found to be in close physical proximity to partially translocated chimeric precursor proteins and is a member of the Com44/Cim44 envelope polypeptides (Wu, C., *et al.* (1994) *J. Biol. Chem. 269*:32264; Ko *et al.* (1995) *J. Biol. Chem. 270*:28601). Introduction and expression of Bce44B was accomplished by subcloning the corresponding cDNA sequence into a T7 promoter-containing plasmid vector (pGEM11Z) and JM109(DE3), a bacterial strain used for overexpressing proteins. See Example 6. Expression of Bce44B in bacteria gave rise to a multiple protein pattern. The largest band corresponded to the full length Bce44B protein with a relative molecular mass of 44 kDa. There were smaller protein bands observed which were most likely derived from a combination of post-translational degradation,

ENSDOCID: <WO 9821346A1 | >

since most of them co-fractionated exclusively with the inclusion bodies and the cytosol, and translation-related events such as internal translation initiations or premature termination of translation. Full length Bce44B co-fractionated primarily with cytoplasmic membranes and crude inclusion bodies. Right-side-out (RSO) and inside-out vesicles (ISO) of bacterial cells were prepared and treated with thermolysin (Kim, Y.J., and Oliver, D.B. (1994) *FEBS Lett.* *339*:175; Kim, Y.J., *et al.* (1994) *Cell* *78*:845) to further determine the nature of the association of Bce44B with the cytoplasmic membranes. See Example 10 through Example 12.

A portion of the membrane-associated Bce44B proteins in both types of vesicles were resistant to post-fractionation protease treatments and became sensitive only in the presence of detergents, *e.g.*, Triton X-100. Vesicles that received a mock protease treatment did not exhibit any Bce44B degradation. The protected portion of the membrane-associated Bce44B proteins is most likely protected from protease degradation by the bacterial membrane, and appears to be associated in a manner similar to its natural environment, the chloroplast envelope. These results indicate that the bacterial protein translocation machinery is capable of recognizing Bce44B and actively integrates it into the cytoplasmic membranes. The membrane-associated Bce44B proteins were present in both the outer membrane and the inner membrane. A large portion of Bce44B fractionated with the outer membrane and a smaller level was found in the inner membrane. The dual nature of the distribution of Bce44B in the bacterial plasma membrane resembles the situation observed in the chloroplast envelope, except that it is opposite to the inner/outer envelope pattern of distribution. The directionality of the Bce44B

distribution pattern may reflect the direction of protein translocation, namely outward in bacteria and inward in chloroplasts.

Determination of the function of Bce44B in bacteria using alkaline phosphatase as a monitor

The directionality of the distribution of Bce44B in the cytoplasmic membrane and the capability of the bacterium to recognize and integrate Bce44B into the membrane raised the possibility that Bce44B may be functional as a foreign component of the bacterial protein transport apparatus. Functionality of Bce44B was therefore assessed by monitoring two transported bacterial proteins, alkaline phosphatase and β-lactamase. Alkaline phosphatase is a single copy gene and is transported into the periplasm upon induction by phosphate starvation (50 μM) (Torriani *et al.* (1960) *Biochim. Biophys. Acta. 38*:460). The growth characteristics of Bce44B-containing bacteria were compared to those of a strain expressing Oee1-Dhfr, a chimeric chloroplast protein precursor. The expression vector used in the Oee1-Dhfr transformed strain was the same as the one in the Bce44B transformed cells. Growth was monitored in M9 minimal media with growth-permissive phosphate levels or with low phosphate levels (50 μM).

The Bce44B-containing cells displayed a growth curve distinct from the Oee1-Dhfr-containing cells. Bce44B cells displayed double peaks during growth in low phosphate, whereas Oee1-Dhfr cells exhibited only one. Bce44B-expressing cells harvested at the time corresponding to the two growth peaks contained a higher level of processed alkaline phosphatase relative to the Oee1-Dhfr-expressing cells. The amount of

processed alkaline phosphatase reflects changes in protein translocation activity. Even though phosphate is a nutrient vital for the survival of the bacterium and although there is most likely a variety of mechanisms operating to allow survival under phosphate starvation, the data show that the presence of Bce44B contributes to enhanced levels of alkaline phosphatase in the periplasm, which in turn allows the cells to grow further by releasing more phosphate from limited sources.

Measurement of protein translocation enhancement using β-lactamase as a transport marker

The enhancement of protein translocation was further investigated using the plasmid-borne multicopy gene, β-lactamase. See Example 14. B-lactamase protein levels are much higher than induced alkaline phosphatase making this system a more sensitive monitor of protein translocation without the complications posed by low phosphate induction. Like alkaline phosphatase, β-lactamase is translocated into the periplasm, where it detoxifies the antibiotic ampicillin, thereby conferring ampicillin-resistance to cells. Because the level of antibiotic resistance conferred also reflects the level of protein transport activity, the level of ampicillin resistance was determined for Bce44B expressing cells and compared to that of Oee1-Dhfr-expressing cells on solid agar media containing increasing concentrations of ampicillin. The plasmid copy numbers were determined to be the same in both strains (approximately 750-800) so that copy number did not contribute to differences in expression of β-lactamase. The Bce44B-expressing cells formed colonies with ampicillin concentrations as high as 3

mg/ml, whereas Oee1-Dhfr-expressing cells could not form colonies beyond 1 mg/ml.

The higher level of antibiotic resistance displayed by the Bce44B-expressing cells was reflected in the enhanced level of transported β-lactamase. Immunoblot analysis of cells grown for four hours in media containing ampicillin concentrations from 50 μg to 3 mg/ml showed that the level of processed β-lactamase is higher on a per cell basis in Bce44B-expressing cells than in Oee1-Dhfr-expressing cells. These data demonstrate that Bce44B enhances the level of β-lactamase translocation, permitting a higher level of antibiotic resistance.

Chemical cross-linking analysis

Direct involvement of Bce44B in the protein translocation process was confirmed by two different methods: chemical cross-linking/co-immunoprecipitation and azide sensitivity. If the integrated form of Bce44B is physically involved in protein translocation, it is probable that the cytoplasmic membrane form of Bce44B is in close physical proximity to translocating β-lactamase. The chemical cross-linking/co-immunoprecipitation results indicated that the two proteins do appear to be in close physical proximity. See Example 16. Bce44B- and Oee1-Dhfr-expressing cells were grown to mid-logarithmic phase and harvested for the preparation of spheroplasts. Washed spheroplasts were subjected to chemical cross-linking with EDAC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide) and the cross-linked complexes were analyzed by immunoblotting with antibodies against Bce44B and β-lactamase. A higher molecular sized cross-linking-generated band of approximately 70 kDa immunoreacted with both antibodies. Bands of this nature were not observed in the

absence of chemical cross-linker or in the Oee1-Dhfr-expressing cells. Cross-linked complexes between β-lactamase and another cytoplasmic membrane protein, Tet, were also not observed in tetracycline-resistant bacterial cells. Tetracycline resistance is due to the presence of a proteinaceous cytoplasmic membrane efflux pump, Tet.

The estimated molecular mass of the cross-linked complex suggests that it likely comprises one each of β-lactamase and Bce44B. These results indicate that β-lactamase is in close physical proximity to Bce44B during translocation, implicating the role of Bce44B in the bacterial protein translocation process.

Sodium azide sensitivity analysis

Sodium azide is a potent inhibitor of bacterial SecA activity, blocking SecA-dependent protein transport in a rapid manner (Oliver *et al.* (1990) *Proc. Natl. Acad. Sci. USA 87*:8227). Both alkaline phosphatase and β-lactamase are known to utilize this pathway for transport. On the other hand, the translocation of proteins across the chloroplast envelope is not affected by this level of azide, even though the translocation of a subset of proteins across the thylakoid membrane is azide-sensitive (Knott and Robinson (1994) *J. Biol. Chem. 269*:7843.). Therefore the Bce44B-expressing cells were tested for azide sensitivity to assess whether protein translocation in these cells has been altered by Bce44B such that it resembles the chloroplastic protein import system. The level of β-lactamase was monitored for 2 hours in the presence of 0.5 mM sodium azide, a concentration that partially inhibits SecA activity in bacteria. Parallel experiments were conducted without sodium azide.

Higher levels of transported β-lactamase were observed in the Bce44B-expressing cells than in the Oee1-Dhfr-containing cells at all time points in the absence of azide. In the presence of sodium azide, the Oee1-Dhfr-expressing cells displayed increasing amounts of the precursor form of β-lactamase and a concomitant decrease in the level of the mature transported form as incubation time progressed. The pattern of β-lactamase accumulation was different in the Bce44B-expressing cells where both precursor and mature forms of β-lactamase increased during the two hours, suggesting that Bce44B can partially compensate for the azide-impaired SecA-dependent protein transport pathway.

Complementation analysis of a temperature-sensitive *secA* bacterial mutant with Bce44B

The effects of Bce44B on β-lactamase transport were further tested in a temperature-sensitive *secA* (*secAts*) mutant bacterial strain (MM52, a gift from Dr. Jon Beckwith, Harvard Medical School, Boston MA) to provide further evidence that the expressed Bce44B component affected protein translocation in bacteria and that the protein transport enhancement was not due solely to the upregulation of the existing bacterial transport machinery such as SecA. See Example 17. The defective protein translocation phenotype in MM52 appears primarily when cells are subjected to temperatures above 30°C. Protein translocation slows down and comes to a halt upon shifting to temperatures above 30°C. This defect is manifested in the gradual accumulation of β-lactamase precursors with a concomitant reduction in the level of the transported mature form of the enzyme. The cause of this defect is due to the temperature-sensitive expression of the main regulatory component of the bacterial protein translocation

machinery, SecA. Effects on protein transport can then be monitored in this strain by tracking the β-lactamase profiles at permissive and non-permissive temperatures.

The same Bce44B gene construct and control Oee1-Dhfr plasmid were introduced into MM52 and its wild type counterpart strain MC4100 (containing the wild-type SecA) via electroporation. The role of Bce44B in protein translocation was assessed as before using three different criteria: 1) the level of antibiotic resistance conferred; 2) the profile of transported versus precursor β-lactamase at permissive and non-permissive temperatures; and 3) the level of sensitivity of protein transport to sodium azide.

The level of antibiotic resistance conferred can reflect the level of protein transport activity; therefore the level of ampicillin resistance was again determined for these strains in the same manner as described above. Each of the strains MM52-Bce44B, MM52-Oee1-Dhfr, MC4100-Bce44B, and MC4100-Oee1-Dhfr were plated out on solid agar media containing increasing concentrations of ampicillin. The plasmid copy numbers were determined to be similar in all strains and were not predicted to contribute to differences in the expression of β-lactamase. MM52-Bce44B-expressing cells were able to form colonies at ampicillin concentrations up to 3 mg/ml whereas MM52-Oee1-Dhfr-expressing cells were not able to form colonies beyond 1 mg/ml ampicillin when grown at 30°C, the permissive temperature. The results for the MC4100-Bce44B and MC4100-Oee1-Dhfr strains were similar to those for the MM52-based strains except that MM52 cells required 48 hours to form colonies versus the 16 hours needed for the MC4100-based strains. The

MM52-based strains were not able to form colonies at 37°C whereas the MC4100-based ones did.

The MM52 strain, although it is capable of growing at 30°C, also showed a reduced level of protein translocation at the permissive temperature and a much lower level at the non-permissive temperature of 37°C. Therefore temperature shift experiments were conducted to assess the effects of Bce44 on protein translocation at permissive and non-permissive temperatures. As described above, the level of processed versus precursor β-lactamase was assessed as a monitor of the ability of Bce44B to enhance the level of bacterial protein translocation. Overnight cultures of MM52-Bce44B, MM52-Oee1-Dhfr, MC4100-Bce44B and MC4100-Oee1-Dhfr cells were used as inocula in parallel experiments. One set of cultures was grown at 30°C throughout the experiment and the other set was shifted from 30°C to 37°C. Samples of cells were harvested at time points representing 0, 1 and 2 hours postshift. Whole cell proteins were subjected to SDS-PAGE and immunoblotting with anti-β-lactamase IgGs. The number of cells assessed was adjusted and normalized before loading the gels. Both MC4100-based strains (MC4100-Bce44B and MC4100-Oee1-Dhfr) showed predominantly the mature form and precursor forms of β-lactamase at both permissive and non-permissive temperatures and at all time points. The MM52-Oee1-Dhfr cells displayed protein translocation deficiencies at both permissive and non-permissive temperatures which were manifested as high ratios of precursor to mature forms of β-lactamase such as that observed in MM52 cells carrying only the vector plasmid. In contrast, the Bce44B-expressing MM52 strain behaved in the same way as the wild-type counterpart, displaying

predominantly the transported mature form of β-lactamase at both permissive and non-permissive temperatures.

Sensitivity to sodium azide was employed to further demonstrate the effects of the Bce44B complementation on protein translocation in the MM52 strain. Overnight cultures of MM52-Bce44B, MM52-Oee1-Dhfr, MC4100-Bce44B, and MC4100-Oee1-Dhfr cells were used as inocula for parallel experiments. These freshly inoculated cultures were grown at 30°C for two hours in media containing increasing concentrations of sodium azide (from 0 to 1 mM). Cells were then harvested, and normalized according to cell number before analysis. The immunoblot results showed that MM52-Oee1-Dhfr-expressing cells displayed protein translocation deficiencies. The precursor form of β-lactamase began appearing at a very low azide concentration (0.125 mM), while MM52-Bce44B-containing cells demonstrated the same behavior as the wild-type counterparts, e.g., MC4100-Bce44B or MC4100-Oee1-Dhfr-expressing cells, where precursor forms of β-lactamase were not detected until the sodium azide concentration reached 0.25-0.375 mM. Azide sensitivity characteristics were further assessed by growing both wild-type and mutant bacterial lines in the media containing 0.5 mM sodium azide for 0, 1 and 2 hr. The immunoblot results showed the same characteristics as in the above experiments, namely that the presence of Bce44B resulted in the same profile of β-lactamase transport as the wild-type cells despite the *secA* mutation.

The complementation results collectively indicate that β-lactamase transport in the *secA* mutant is influenced by Bce44B directly in enhancing transport rather than by upregulation of the bacterial protein translocation machinery.

Analysis of protein translocation using deletion constructs of Bce44B

Two deletion mutants of Bce44B were made by removing the N-terminal 42 amino acids and by removing 17 amino acids at the carboxyl end of the N-terminal 42 amino acid region. These are designated K117 (Figure 4) and K118 (Figure 5), respectively. The two deletion constructs of Bce44B were subjected to the same set of analyses as outlined above to assess the effects of the deletions on the protein translocation enhancement property of Bce44B.

The deletion constructs were introduced into JM109(DE3) as described in Example 7, using the same T7 promoter-bearing plasmid vector to facilitate expression. The immunoblotting results indicate that both deletions affected the expression level of Bce44B in *E. coli*. The lowest levels were exhibited by cells harboring K118. Cells harboring K117 also demonstrated lower levels compared to unaltered Bce44B-containing cells but substantially higher levels than K118. The level of mature β-lactamase in K117-harboring cells was lower than unaltered Bce44B-containing cells on a per cell basis. The β-lactamase levels were the lowest in the K118-containing cells. The lower level of translocated β-lactamase was reflected in the level of ampicillin concentration tolerated by the corresponding strain. In contrast to the Bce44B-containing cells, K117 and K118 exhibit differences in the ability to form colonies on plates with increasing concentrations of ampicillin. K117 can only form colonies at 500 μg/ml ampicillin and K118 can only form colonies up to 250 μg/ml ampicillin.

Wild-type Bce44B is targeted to both outer and inner membranes. The amount of Bce44B present in the outer membrane is

SUBSTITUTE SHEET (RULE 26) WO 98/21346A1

higher than in the inner membrane. In comparison to Bce44B, K117 is still targeted to both inner and outer membranes despite the deletion. Distribution of the truncated K117 protein occurred in the same fashion as for the wild-type Bce44B protein, albeit at significantly lower amounts. The K118 deletion protein is targeted only to the inner membrane in even lower amounts than either Bce44B or K117. Both K117 and K118 displayed resistance to thermolysin to the same degree as unaltered Bce44B and were completely sensitive only in the presence of 0.1% Triton X-100. In conclusion, the two deletions represented by K117 and K118 appear to affect the function of Bce44B in protein translocation.

Azide sensitivity experiments were conducted on the K117- and K118-containing cells and compared to Bce44B- and Oee1-Dhfr-expressing cells using liquid cultures containing increasing concentrations of sodium azide (0 mM, 0.125 mM, 0.25 mM, 0.375 mM, 0.5 mM, 0.625 mM, 0.75 mM, 0.875 mM, 1 mM) for two hours. The results show increasing amounts of the precursor form with a concomitant decrease in the amount of the mature form of β-lactamase in all four types of cells. The azide inhibition of protein translocation patterns for K117 and K118 was similar to the inhibition pattern for Oee1-Dhfr-containing cells.

The results collectively suggest that the deletions affected the protein transport enhancing ability of Bce44B protein. These Bce44B mutants are good candidates for use in a dominant-negative approach to reduce the activity of a naturally-occurring Bce44B protein.

Technological applications

Bce44B is a member of the Com44/Cim44 chloroplast envelope proteins. This invention provides, for the first time, evidence

BNSDOCID: <WO 9821346A1 >

that Bce44B is part of the chloroplast protein import apparatus and can transport substances across plant membranes. One surprising and important advantage of Bce44B protein is its unique capacity to effectively transport substances by itself without requiring a multicomponent transport assembly which would be difficult to incorporate successfully into other organisms. As evidence of this advantage, data is provided demonstrating that Bce44B can translocate substances across bacterial (prokaryotic) membranes as well as plastid membranes. Both systems recognize and incorporate Bce44B in the same manner relative to the direction of protein translocation.

As in the chloroplast envelope, translocating precursor proteins are found in close physical proximity to Bce44B in the bacterial membrane. The integrated protein in the bacterial membrane additionally affects the SecA-dependent protein export of alkaline phosphatase and β-lactamase by enhancing their levels via a less azide-sensitive mode of action. It is not known if Bce44B acts independently by forming new more azide-tolerant pathways, most likely via the assembly of chimeric protein translocation machinery, or by altering a portion of the pathways already in existence, thus becoming less azide-sensitive overall (similar to the systems found in chloroplasts). The combination of the chemical cross-linking results and the partial nature of the azide insensitivity tends to point to the first possibility.

This invention also provides evidence of parallelism of protein translocation in bacteria and plastids, demonstrating that there are common features in protein translocation systems among different organisms and providing, as a result, methods of translocating substances for any organism. Thus, the methods of this invention provide a novel and powerful way to elucidate the

roles and functions of components of the plastid protein transport machinery by utilizing bacteria, and to further allow access to and application of all of the molecular tools available for studying various aspects of bacteria for studies of plastid protein import.

Since the mitochondrion is thought to have evolved in a similar fashion to plastids via an endosymbiotic mode, the bacterial approach described for studying protein uptake into plastids can also be utilized for constructing methods to alter import of substances in mitochondria. The mode of protein importation in mitochondria and the structure of the organelle closely resemble those found in plastids; thus, those of skill in the art will recognize that minor modification is required to apply the methods described in this invention for use in mitochondria and other double membrane compartments.

In another aspect, this invention relates to antibodies which bind the polypeptides described herein. Such antibodies can be used to locate sites of translocation or membrane integration activity in cells. Fusion proteins comprising Bce44B and an additional peptide, such as a protein tag, can also be used to detect sites of Bce44B activity in cells of prokaryotes and eukaryotes. Detection of sites of activity is useful to help understand the structure and function of membranes, especially those of plastids and mitochondria. These antibodies can also be useful to inhibit translocation of substances across membranes or integration of substances into membranes.

Isolated DNA is introduced into plant cells of a target plant by well-known methods, such as *Agrobacterium*-mediated transformation, microprojectile bombardment, microinjection or electroporation. Cells carrying the introduced isolated and/or

BNSDOCID <WO 9821346A1 | >

recombinant DNA can be used to regenerate transgenic plants which have altered phenotypes, therefore becoming sources of additional plants either through seed production or non-seed asexual reproductive means.

The methods of this invention can be used to provide plants, seeds, plant tissue culture, plant parts, cells, and protoplasts containing one or more nucleic acids which comprise a modified or isolated introduced gene encoding a Bce44B protein or its functional equivalent which alters transport across a cellular membrane or incorporation of molecules into membranes. Plants parts can include roots, leaves, stems, flowers, fruits, meristems, epicotyls, hypocotyls, cotyledons, pollen and embryos.

The present invention also relates to transgenic plants, or cells or tissues derived from such plants, in which membrane transport or incorporation of molecules into membranes is altered directly or indirectly through application of the methods of this invention. The term "transgenic plants" includes plants or photosynthetic protists which contain introduced DNA which, if transcribed and translated, changes the amount or type of one or more plant products compared to a wildtype (naturally-occurring) plant of the same species or variety grown under the same conditions. Transgenic plants include those into which isolated and/or recombinant nucleic acids have been stably inserted and their descendants, produced from seed, vegetative propagation, cell, tissue or protoplast culture, or the like wherein such alteration is maintained. The introduced DNA which is originally inserted into the plants or plant cells or protoplasts can include additional copies of genes found in the naturally-occurring organism.

In one embodiment, the methods of this invention can be used to stably transform the genome of plastids or mitochondria. Whether through nuclear, plastid, or mitochondrial constructs, the ability to alter plastid or mitochondrial components is a great advantage. Especially in plants, the alteration can be contained because the plastids of most plants are maternally inherited, thus the altered genes will not be transmitted in the pollen which is freely disseminated. The risk of transmission to wild-type (native) plants is then greatly reduced. Further, disruption of mitochondria, especially in pollen or pollen tube formation can be desirable for transgenic containment when the disruptive peptides or polypeptides are expressed in a tissue-specific manner and/or during a particular stage of development. Methods of producing pistil-specific or anther-specific expression of a nucleotide sequence to produce either female or male sterility, respectively, can be found, for example, in Nasrallah, *et al.* (1994) PCT/US94/04557 (WO 94/25613).

Those of skill in the art will recognize the methods of this invention for enhancing import or export of naturally-occurring plant products such as proteins, oils, carbohydrates, and combinations thereof into and out of organelles or into or out of cells has the possibility of an almost infinite number of applications. Besides enhancing accumulation or export of naturally-occurring products, the transgenic plants can also contain introduced genes which encode useful products whose accumulation or harvest is facilitated by enhanced membrane transport. In particular, antigens for vaccine purposes, antibodies, blood products, enzymes and the like, as well as insect or disease inhibitors for plants are examples of products which can be provided from other sources including mammals.

For example, transgenic plants expressing foreign peptides, such as the binding subunit of *E. coli* heat-labile enterotoxin have been shown to accumulate these antigens in leaves and tubers (Haq *et al.* (1995) *Science 268*:714-716). Further, transgenic plant cells have been shown to express and assemble secretory antibodies (Ma, *et al.* (1995) *Science 268*:716). Enhanced membrane transport as described in the present invention can be used to increase the solubilization and accumulation of such plant-derived products in the edible portions of plants or in the portions of plants intended to be harvested for extraction of these compounds, or even to change the size of plant parts. Transport can be altered, for example, in any plant organ; *i.e.*, stems, roots, leaves, flowers and fruits.

The methods provided herein can enhance the accumulation and/or transport of substances for plant disease resistance. For example, fungal pathogens (*e.g.*, *Sclerotinia* sp.) produce oxalic acid which helps to break down the cell walls of plants and promotes fungal growth. Isolated DNA encoding oxalate decarboxylase or oxalate oxidase can be inserted into plants wherein the expressed enzyme can be solubilized by Bce44B protein and its accumulation can be directed through Bce44B-assisted transport to cell walls and intercellular spaces where it can degrade and detoxify oxalic acid, thereby protecting plants against this pathogen. See, for example, European Patent Applications EP 673,416 (September 27, 1995) and EP 531,498 (March 17, 1993). These enzymes can also be used to reduce the oxalic acid content of plants in which the oxalic acid content is high and results in toxicity of the plant or plant part when ingested (e.g., rhubarb).

Those of skill in the art can understand that the variety of transgenic products which can be produced and translocated in prokaryotic and eukaryotic organisms by the methods described herein is broad and encompasses many important naturally-occurring and foreign substances which are regulatory or are products themselves. These include, for example, storage products such as sugars, starches, pigments, and the like. If naturally-occurring in the photosynthetic organism, the product may be produced at higher levels, compartmentalized in a different part of the cell, such as the plastid, mitochondrion, or vacuole, or even in a different organ, such as the flower, seed, root or leaf. Thus, pigments can be expressed and/or accumulated at higher levels to enhance the color of the plant or plant part normally producing the pigment by methods provided herein. Further, through the same methods, a novel color can be produced in a plant or imparted to a plant organ by linking the gene encoding the pigment (or the proteins which catalyze pigment synthesis) to the constructs described herein so that the pigment gene products are expressed and translocated.

Alternatively, this invention provides methods for varying the phenotype of seeds and other storage organs of plants. These novel products or combination of products can be provided by enhancing the translocation of molecules to be stored or by modifying the composition of cellular membranes to alter the translocation of particular products. Thus, in addition to increasing the overall amount of stored substances, thus increasing the nutritive value of the seed, alterations can include modifying the fatty acid composition in seeds by changing the ratio and/or amounts of the various fatty acids as they are produced. Alternatively, improvements in the amino acid

composition of storage proteins can be generated. Of particular interest as target substances are the storage proteins of seeds, such as napin, cruciferin, β-conglycinin, phaseolin, brazil nut protein, other 2S or 7S proteins, or the like, as well as proteins involved in fatty acid biosynthesis, such as acyl carrier protein.

When useful proteins are expressed at high levels in the transgenic plants (or other eukaryotes of this invention), these levels could be toxic to the cell or organism. Therefore, it can be important to sequester such highly expressed proteins in compartments such as plastids, especially plastids of storage tissues. As an example, genetically-engineered plants which express Bt (*B. thuringiensis*) toxins as recombinant proteins may show inhibited growth due to high levels of the expressed toxins, which are intended to provide insect resistance by poisoning grazing insects. However, increased levels of these proteins will be required as endemic populations of insects develop resistance to the presently expressed levels in recombinant plant tissues. Sequestering this protein in the plastids following expression either within or outside of the plastid, could reduce its toxic effects in the cell. Vacuoles can also be used to sequester toxic compounds.

Enhanced expression and translocation of substances into plastid compartments or other organelles can facilitate a preferred or more efficient method of purifying or processing protein products or other products generated from a transgenic plant. Further, enhancing translocation of a transgenic product into plastids can provide a separate environment in which high concentrations of proteins can induce an "inclusion bodies" effect similar to that commonly observed in bacterial

SUBSTITUTE SHEET (RULE 26)

overexpression, thus facilitating isolation of the preferred product.

To produce transgenic plants of this invention, a construct comprising the gene for Bce44B or nucleic acid encoding its functional equivalent and a promoter are incorporated into a vector as described in Example 19 or through other methods known and used by those of skill in the art. The construct can also include any other necessary regulators such as terminators or the like, operably linked to the coding sequence. It can also be beneficial to include a 5' leader sequence, such as the untranslated leader from the coat protein mRNA of alfalfa mosaic virus (Jobling, S.A. and Gehrke, L. (1987) *Nature 325*:622-625) or the maize chlorotic mottle virus (MCMV) leader (Lommel, S.A., *et al.* (1991) *Virology 81*:382-385). Those of skill in the art will recognize the applicability of other leader sequences for various purposes.

Targeting sequences are also useful and can be incorporated into the constructs of this invention. A targeting sequence is usually translated into a peptide which directs the polypeptide product of the coding nucleic acid sequence to a desired location within the cell, such as to the plastid, and becomes separated from the peptide after transit of the peptide is complete or concurrently with transit. Examples of targeting sequences useful in this invention include, but are not limited to, the yeast mitochondrial presequence (Schmitz, *et al.* (1989) *Plant Cell 1*:783-791), the targeting sequence from the pathogenesis-related gene (PR-1) of tobacco (Cornellisen, *et al.* (1986) *EMBO J.* 5:37-40), vacuole targeting signals (Chrispeels, M.J. and Raikhel, N.V. (1992) *Cell 68*:613-616), secretory pathway sequences such as those of the ER or Golgi (Chrispeels, M.J.

BNSDOCID: <WO 9821346A1 I_>

(1991) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 42:21-53). Intraorganellar sequences may also be useful for internal sites, *e.g.*, thylakoids in chloroplasts. Theg, S.M. and Scott, S.V. (1993) *Trends in Cell Biol.* 3:186-190.

In addition to 5' leader sequences, terminator sequences are usually incorporated into the construct. In plant constructs, a 3' untranslated region (3' UTR) is generally part of the expression plasmid and contains a polyA termination sequence. The termination region which is employed will generally be one of convenience, since termination regions appear to be relatively interchangeable. The octopine synthase and nopaline synthase termination regions, derived from the Ti-plasmid of *A. tumefaciens*, are suitable for such use in the constructs of this invention.

Any suitable technique can be used to introduce the nucleic acids and constructs of this invention to produce transgenic plants with an altered genome. For grasses such as maize, microprojectile bombardment (see for example, Sanford, J.C., *et al.*, U.S. Patent No. 5,100,792 (1992) can be used. In this embodiment, a nucleotide construct or a vector containing the construct is coated onto small particles which are then introduced into the targeted tissue (cells) via high velocity ballistic penetration. The vector can be any vector which permits the expression of the exogenous DNA in plant cells into which the vector is introduced. The transformed cells are then cultivated under conditions appropriate for the regeneration of plants, resulting in production of transgenic plants. Transgenic plants carrying the construct are examined for the desired phenotype using a variety of methods including but not limited to an appropriate phenotypic marker, such as antibiotic resistance

or herbicide resistance, or visual observation of the time of floral induction compared to naturally-occurring plants.

Other known methods of inserting nucleic acid constructs into plants include Agrobacterium-mediated transformation (see for example Smith, R.H., *et al.*, U.S. Patent No. 5,164,310 (1992)), electroporation (see for example, Calvin, N., U.S. Patent No. 5,098,843 (1992)), introduction using laser beams (see for example, Kasuya, T., *et al.*, U.S. Patent No. 5,013,660 (1991)) or introduction using agents such as polyethylene glycol (see for example Golds, T. *et al.* (1993) *Biotechnology*, 11:95-97), and the like. In general, plant cells may be transformed with a variety of vectors, such as viral, episomal vectors, Ti plasmid vectors and the like, in accordance with well known procedures. The method of introduction of the nucleic acid into the plant cell is not critical to this invention.

The methods of this invention can be used with *in planta* or seed transformation techniques which do not require culture or regeneration. Examples of these techniques are described in Bechtold, N., *et al.* (1993) *CR Acad. Sci. Paris/Life Sciences 316*:118-93; Chang, S.S., *et al.* (1990) *Abstracts of the Fourth International Conference on Arabidopsis Research*, Vienna, p. 28; Feldmann, K.A. and Marks, D.M (1987) *Mol. Gen. Genet. 208*:1-9; Ledoux, L., *et al.* (1985) *Arabidopsis Inf. Serv. 22*:1-11; Feldmann, K.A. (1992) *In*: Methods in *Arabidopsis* Research (Eds. Koncz, C., Chua, N-H, Schell, J.) pp. 274-289; Chee, *et al.*, U.S. patent, Serial No. 5,376,543.

The transcriptional initiation region may provide for constitutive expression or regulated expression. Many promoters are available which are functional in plants. The term "promoter" refers to a sequence of DNA, usually upstream (5') of

the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription.

Constitutive promoters for plant gene expression include, but are not limited to, the octopine synthase, nopaline synthase, or mannopine synthase promoters from *Agrobacterium*, the cauliflower mosaic virus (35S) promoter, the figwort mosaic virus (FMV) promoter, and the tobacco mosaic virus (TMV) promoter. Constitutive gene expression in plants can also be provided by the glutamine synthase promoter (Edwards, *et al.* (1990) *PNAS* 87:3459-3463), the maize sucrose synthetase 1 promoter (Yang, *et al.* (1990) *PNAS* 87:4144-4148), the promoter from the Rol-C gene of the TLDNA of Ri plasmid (Sagaya, *et al.* (1989) *Plant Cell Physiol.* 30:649-654), and the phloem-specific region of the pRVC-S-3A promoter (Aoyagi, *et al.* (1988) *Mol. Gen. Genet.* 213:179-185).

Heat-shock promoters, the ribulose-1,6-bisphosphate (RUBP) carboxylase small subunit (ssu) promoter, tissue specific promoters, and the like can be used for regulated expression of plant genes. Developmentally-regulated, stress-induced, wound-induced or pathogen-induced promoters are also useful.

The regulatory region may be responsive to a physical stimulus, such as light, as with the RUBP carboxylase ssu promoter, differentiation signals, or metabolites. The time and level of expression of the sense or antisense orientation can have a definite effect on the phenotype produced. Therefore, the promoters chosen, coupled with the orientation of the exogenous DNA, and site of integration of a vector in the genome, will determine the effect of the introduced gene.

Specific examples of regulated promoters also include, but are not limited to, the low temperature Kin1 and cor6.6 promoters (Wang, *et al.* (1995) *Plant Mol. Biol. 28*:605; Wang, *et al.* (1995) *Plant Mol. Biol. 28*:619-634), the ABA inducible promoter (Marcotte Jr., *et al.* (1989) *Plant Cell 1*:969-976), heat shock promoters, such as the inducible hsp70 heat shock promoter of *Drosphilia melanogaster* (Freeling, M., *et al.* (1985) *Ann. Rev. of Genetics 19*: 297-323), the cold inducible promoter from *B. napus* (White, T.C., *et al.* (1994) *Plant Physiol. 106*:917), the alcohol dehydrogenase promoter which is induced by ethanol (Nagao, R.T., *et al.*, Miflin, B.J., Ed. *Oxford Surveys of Plant Molecular and Cell Biology*, Vol. 3, p 384-438, Oxford University Press, Oxford 1986), the phloem-specific sucrose synthase ASUS1 promoter from *Arabidopsis* (Martin, *et al.* (1993) *Plant J. 4*:367-377), the ACS1 promoter (Rodrigues-Pousada, *et al.* (1993) *Plant Cell* 5:897-911), the 22 kDa zein protein promoter from maize (Unger, *et al.* (1993) *Plant Cell* 5:831-841), the ps1 lectin promoter of pea (de Pater, *et al.* (1993) *Plant Cell* 5:877-886), the phas promoter from *Phaseolus vulgaris* (Frisch, *et al.* (1995) *Plant J.* 7:503-512), the lea promoter (Thomas, T.L. (1993) *Plant Cell 5*:1401-1410), the E8 gene promoter from tomato (Cordes, *et al.* (1989) *Plant Cell 1*:1025-1034), the PCNA promoter (Kosugi, *et al.* (1995) *Plant J.* 7:877-886), the NTP303 promoter (Weterings, *et al.* (1995) *Plant J.* 8:55-63), the OSEM promoter (Hattori, *et al.* (1995) *Plant J.* 7:913-925), the ADP GP promoter from potato (Muller-Rober, *et al.* (1994) *Plant Cell 6*:601-604), the Myb promoter from barley (Wissenbach, *et al.* (1993) *Plant J.* 4:411-422), and the plastocyanin promoter from *Arabidopsis* (Vorst, *et al.* (1993) *Plant J.* 4:933-945).

Transgenic plants of this invention can contain isolated or recombinant nucleic acids which preferentially modify plastid transport pathways which are present in green tissues, or which are present in actively growing tissues or in storage tissues or organs such as seeds. The plastids of plant tissues are desirable targets for modifications to provide increased photosynthetic capacity or to provide mechanisms for disease or stress resistance. Alternatively, transgenic plants can contain introduced isolated or recombinant nucleic acids which alter the transport capability of other organelles or of the plasma membrane of cells. In this manner, different products can be accumulated, exported or imported to modify the capability of the plant to express and localize one or more products compared to the expression and accumulation of the same product(s) in a plant of the same variety without said introduced isolated or recombinant nucleic acids when grown under identical conditions.

Further, this invention includes a method of producing a transgenic plant containing, in addition to isolated nucleic acids which encode a Bce44B polypeptide or its functional equivalent so that membrane transport is altered, at least one nucleic acid which encodes a polypeptide for production of a useful foreign product. Coupled with the altered membrane transport system in the cells of the plant, it is possible to design a plant wherein, when all of the inserted nucleic acids are expressed, the result is the large scale and inexpensive production of valuable proteins or other products in a particular plant tissue or at a particular stage of development.

The methods described herein can be applied to all types of plants and other photosynthetic organisms, including: angiosperms (monocots and dicots), gymnosperms, spore-bearing or

vegetatively-reproducing plants and the algae (including the blue-green algae). Further, the methods of this invention are suited to enhance translocation of substances in all prokaryotes. It is understood that prokaryotic organisms lack plastids and other organelles, but that the photosynthetic membranes or cell membranes of these organisms can be modified to alter photosynthetic capacity and products and/or translocation of other products through expression of a Bce44B polypeptide or its functional equivalent and incorporation of the same in the appropriate cellular membrane. It is also likely that the methods described herein can be applied, without undue experimentation, to enhance transport of substances in nonphotosynthetic eukaryotes such as the fungi and Animalia, particularly in mitochondria, due to their analogous evolutionary history with plastids.

Transgenic plants containing the constructs described herein can be regenerated from transformed or transfected cells, tissues or portions of plants by methods known to those of skill in the art. A portion of a plant is meant to include any part capable of producing a regenerated plant. Thus, this invention encompasses a cell or cells, tissue (especially meristematic and/or embryonic tissue), protoplasts, epicotyls, hypocotyls, cotyledons, cotyledonary nodes, pollen, ovules, stems, roots, leaves, and the like. Plants may also be regenerated from explants. Methods will vary according to the plant species.

Seed can be obtained from the regenerated plant or from a cross between the regenerated plant and a suitable plant of the same species. Alternatively, the plant may be vegetatively propagated by culturing plant parts under conditions suitable for the regeneration of such plant parts. For example, plants can be

regenerated from cultured pollen, protoplasts, meristems, hypotcotyls, epicotyls, stems, leaves, and the like.

Isolated and purified Bce44B protein or polypeptides, and epitopic fragments thereof, can also be used to prepare antibodies which can be used, for example, for localization of sites of Bce44B function and to analyze developmental pathways in plants. Antibodies that specifically bind a Bce44B protein can detect Bce44B expression in specific cells or tissues of plants. This information can be used to determine how and when Bce44B acts to transport substances across membranes.

Antibodies used in the methods of the invention can be polyclonal, monoclonal, or antibody fragments, and the term antibody is intended to encompass polyclonal antibodies, monoclonal antibodies and antibody fragments. Antibodies of this invention can be raised against isolated or recombinant Bce44B proteins or polypeptides. (See, *e.g.*, Examples 21-23). Preparation of immunizing antigen, and antibody production can be performed using any suitable technique. A variety of methods have been described (see e.g., Harlow, E. and D. Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; Ausubel *et al.* (1994) *Current Protocols in Molecular Biology*, Vol. 2, Chapter 11 (Suppl. 27) John Wiley & Sons: New York, NY).

Antibodies used in the methods of this invention can be labeled or a second antibody that binds to the first antibody can be labeled by some physical or chemical means. The label can be an enzyme which is assayed by the addition of a substrate which upon reaction releases an ultraviolet or visible light-absorbing product or it can be a radioactive substance, a chromophore, or a fluorochrome. E. Harlow and D. Lane (1988) *supra*.

The Bce44B protein component works closely with the 97 kDa component of the plastid protein import machinery. The stoichiometry of Bce44B relative to other components of the plastid protein transport machinery, such as the 97 kDa (Cim97) component, appears to be a key operating factor in the import mechanism. Stoichiometry differences can provide a method to detect changes in import activity in plant tissues. Differences in stoichiometry, such as a higher ratio of Bce44B to Cim97, can be used, for example, to detect plants that will give rise to greater biomass and larger seeds. Larger seeds correlate with alterations to the stoichiometry of the components of plastid protein import machinery in naturally occurring plants and even between different cultivars. These values can be detected by the antibodies disclosed herein. This approach can also be developed into a diagnostic kit which can be used to determine whether a plant will yield, for example, larger seeds. A diagnostic kit for this purpose can be valuable tool by which plant breeders can screen for desirable products of breeding programs before maturation. There are many advantages of having an early measurement from such a kit. Most important, it would eliminate the laborious task of dealing with all products of breeding and allow an early focus on the most promising plants in a population.

The observed differences in stoichiometry of Bce44B relative to other components of the plastid protein transport machinery also suggest that the action of Bce44B can be regulated or affected by the introduction of other components of the plastid protein transport machinery, *e.g.*, Cim97. Thus, it is possible to alter or fine tune the activity of the Bce44B gene by introducing regulators to the DNA constructs of this invention.

Selected effects in the resulting transgenic plants can be achieved by manipulation of the components to a relationship that provides the desired plant traits, *e.g.*, seed size.

In another method of this invention, Bce44B polypeptide or its functional equivalent can be used to detect and analyze protein/protein transport interactions. Fusion proteins for this purpose can be prepared by fusing Bce44B nucleic acid encoding a Bce44B polypeptide or its functional equivalent with heterologous DNA encoding a different polypeptide (one not related or homologous to the Bce44B polypeptide), such as a protein tag. The resulting fusion protein can be prepared in a prokaryotic cell (e.g. *E. coli*), isolated, labeled and used essentially like antibodies to detect binding sites of Bce44B/protein interactions. See Ron and Dressler (1992) *Biotech 13*:866-869; Smith and Johnson (1988) *Gene 67*:31-40.

Those skilled in the art can recognize the advantages of using selective promoters as well as constitutive promoters in the plasmids and vectors of this invention. Selective promoters can be used to alter the transport of substances during times of stress or at particular growth stages. For example, increased photosynthesis through increased import of precursor materials such as chlorophyll binding proteins into the plastid can boost the growth of a seedling so that it is well established and larger than a naturally-occurring plant when periods of stress occur. Thus, it can have a deeper, more extensive root system which gives it more access to water and helps it survive droughts.

Larger seeds produced by transgenic plants can result in more product and better germination. These seeds can also be planted deeper since the increased reserve metabolites will allow

them to grow more quickly (or over a longer period of time), if necessary, before the seedling reaches the surface of the soil. Increased transport can result in larger roots (*e.g.*, carrots) or in larger tubers (*e.g.*, potatoes).

Enhanced protein transport through incorporation of the Bce44B gene can increase the overall vegetative size of plants, thus increasing yields. This growth enhancement is probably due to the plant's need to regulate internal stress relating to protein concentration and to the osmotic preference of cellular compartments such as the plastid, mitochondrion or vacuole. Manipulation of protein concentration through alteration of protein transport may disturb osmotic pressures in such compartments. For example, to compensate for higher osmotic pressure in a plastid (due to increased protein import into the plastid), the cell could signal division of the plastid. If this division were not adequate to restore normal pressure, cell division could result. Eventually, these effects would produce faster growth and larger plants. In accordance with this model, incorporation and expression of the Bce44B gene in plants is a method of eliciting faster growth and producing larger plants.

Further, antisense or dominant-negative constructs can be used to inhibit transport of substances, thereby accumulating substances at selected sites. Since Bce44B is a general component of a plant system, site-specific or inducible promoters can be used to prevent lethal effects due to complete inhibition.

In another aspect, the methods of this invention can provide means for the selection of cells and organisms which have been infected, transfected or transformed. Through these means, selectable markers are available without concern for the toxicity of the products, their effect on the environment, or concern that they will have to be regulated. At present, selection methods for plants, plant cells or tissues which have successfully incorporated inserted nucleic acid constructs usually involve detection of gene products of marker genes inserted as part of the nucleic acid construct. The most common strategy for obtaining transgenic cells has been the inclusion of drug-resistance determinants, such as kanamycin resistance, in the vector inserted into the cells. Although addition of drugs to growth medium allows selection for cells containing the incorporated nucleic acids, addition of antibiotics is unacceptable in many instances because of cost and possible contamination of the end product.

Inclusion of additional Bce44B protein into plastids results in enhanced transport of materials into plastids. This in turn leads to increased photosynthetic end products such as carbohydrates and lipids which are stored in seeds. The seeds of such transgenic plants are larger than the wildtype variety (see Example 20) thereby providing a marker which can be used reliably to circumvent the inclusion of antibiotics to the culture medium. Therefore, a method of detecting transformation, infection or transfection in a cell comprises: incorporating isolated nucleic acid encoding Bce44B or its functional equivalent into the cell; maintaining the cell or organism under conditions appropriate for expression of Bce44B or its functional equivalent; and then detecting the enhanced transport in the cell, wherein detection

of enhanced transport is indicative of a transformed, infected or transfected cell. Using this method, a plant derived from a transformed, infected or transfected cell produces an enlarged seed compared to a seed of an untransformed, infected, or transfected plant of the same variety under the same growing conditions, and the enlarged seed provides a selectable marker.

In one particular embodiment, the constructs of this invention can be used as a selectable marker and means of removing undesirable insertions of Ti plasmid DNA to avoid potential regulatory problems with transgenic plants and plant products. During Agrobacterium-mediated transformation, part of the Ti plasmid is incorporated in the plant genome. The DNA encoding Bce44B can be incorporated into another part of a typical Agrobacterium binary vector, outside the T-DNA region, and placed under the control of an appropriate promoter for expression in plants to allow detection of transformation. Thus, the Ti plasmid portion is "tagged" with this gene, thereby producing a visual marker which segregates with the Ti plasmid portion. Detection of transformation can be determined by examining protein import modifications or visually by examining transgenic seed size and other phenotypic characteristics of increased transport. The Ti plasmid can then be selected out through subsequent generations by appearance of increased transport (*i.e.*, larger plants, etc.) or by larger seeds resulting from expression of the Bce44B DNA. Normally-sized seeds would indicate the loss of the Bce44B-linked Ti plasmid but presumably still contain the transgene of choice. Other undesirable genes can be expunged by this method. The resulting transgenic plants can then be transformed again via the same approach for many cycles. Each cycle will result in the addition

of a new transgene but remove the Bce44B marker gene and the accompanying Ti plasmid DNA sequence. Thus, this method permits the re-use of the selectable marker "Bce44B" within one or two generations, which is currently difficult to achieve for multigenic manipulations.

Likewise, increased photosynthetic capability leads to faster growth of cells, tissue culture, and the individual plants compared to wildtype materials, thus providing other means to identify successful incorporation of transport protein nucleic acids. Thus, increased transport can potentially bypass culturing as commonly practiced.

Increased foreign product can also constitute a marker which allows for selection of cells or organisms having incorporated nucleic acids encoding both the Bce44B protein or its functional equivalent and one or more foreign products. Peptides, phosphatases, insulin, antibodies, or even livestock hormones can be used as such markers. For example, a bacterial cell producing human growth hormone (hGH) will show an increase in translocation of hGH if Bce44B or an equivalent polypeptide is incorporated into the bacterial cell membranes. Such substances can be detected by immunodetection, enzymatic assay or the like.

Thus, one method of detecting transformation, infection or transfection in a cell comprises; incorporating both isolated nucleic acid encoding Bce44B or its functional equivalent and another gene encoding a product into a cell or organism; maintaining the cell or organism under conditions appropriate for expression of Bce44B or its functional equivalent and the product encoded by the other gene; and then detecting the enhanced transport of the product or an endogenous polypeptide, wherein

detection of enhanced transport is indicative of a transformed, infected or transfected cell.

Increased resistance to ampicillin or increased ability to grow under conditions of phosphate starvation provide methods by which transgenic cells carrying Bce44B and an additional gene for a product can be selected from untransformed cells. For example, compounds such as alkaline phosphatase can provide the basis for a complementation system for the selection and maintenance of such product genes in bacterial or yeast hosts. Genes encoding Bce44B protein (or its functional equivalent), and a commercially-important product can be incorporated into plasmids and the plasmids inserted into host cells which are maintained in environments requiring additional translocation of alkaline phosphatase by the cells. Through complementation of transport genes of the host, which defect cannot be overcome without additional translocation of alkaline phosphatase or nutritional supplementation, the plasmids are stabilized in the host cells. Thus, plasmids carrying genes for commercially-important products can be maintained in the cells with the increased translocation of alkaline phosphatase providing the stabilization of the product-producing cells. A similar strategy is possible in plants and would most likely involve complementation of plastid or mitochondrial genes.

Those of skill in the art will recognize the advantages of being able to alter translocation processes in bacteria, such as the secretion or import of substances, and even to influence expression and solubilization of products produced in these prokaryotic cells. The method of increasing translocation of fermentation products for harvesting through the activity of

BNSDOCID: <WO 9821346A1 | >

Bce44B or its functional equivalent is an extremely valuable and useful mechanism by itself.

There are an unlimited number of products which either isolated from or harvested as part of the organism producing them would be useful. In addition, strains of bacteria harboring genes encoding proteins which enhance translocation (and therefore, utilization) of substances can be used to innoculate the intestines of cows and other ruminants which depend on microbes for their digestive processes.

Bacterial cells can easily be grown to high densities, which is the basis for the inexpensive, high-yielding fermentation processes developed for the production of protein products on a large industrial scale. However, the production and transport of foreign proteins in bacteria do not always reach optimal and economically viable levels, especially when fusion proteins are overproduced. The levels expressed and translocated depend largely on the fusion protein in question and the bacterium's ability to efficiently recognize, transport and process the particular foreign protein. Although translocation is achieved in principle by attaching an N-terminal signal peptide to the desired protein, the transported levels achieved relative to the levels of protein expressed are not proportionally high in most cases. The translocation of foreign proteins into the media or periplasm is desirable to reduce processing cost. Translocation which results in secretion or exportation enhances the purity of the desired overexpressed product by physically partitioning the expressed product from cytoplasmic proteins. Such translocation avoids cytoplasmic toxicity, attack by endogenous proteases and N-terminal methionine extension. Furthermore, the accumulation of products in a more oxidizing environment where disulphide bond

formation may proceed is highly desirable, so that the protein may fold into a soluble, biologically active conformation (Hockney (1994) *TIBS 12*:456; Missiakas *et al.* (1993) *Proc. Natl. Acad. Sci. USA 90*:7084).

The *E. coli* translocation machinery does not always work efficiently with overexpressed proteins. Although there have been relatively successful approaches for optimizing protein translocation, each strategy has limitations and the positive enhancements appear restrictive for general application, *i.e.*, not every protein works. One strategy is to engineer the signal peptide to make the protein a better substrate for the transport machinery (Klein *et al.* (1992) *Protein Eng. 5*:511). Another approach is to elevate components of the endogenous transport apparatus and this has met with some success for two fusion proteins, OmpA-human interleukin-6 and human granulocyte-colony stimulating factor (Perez-Perez, *et al.* (1994) *Biotechnology 12*:178). However, as the authors in this latter study point out, the levels are still well below 100 percent.

There are also other means of enhancing translocation of heterologous proteins in *E. coli* via strategies that alter the efficiency of protein synthesis. Translational level strength is critical to the secretion of heterologous proteins in these bacteria. Simmons, L.C. and Ynasura, D.G. (1996) *Nature Biotechnology 14*:629-634. Thus translational efficiency can be improved by manipulating 5' untranslated region sequences.

Another strategy can involve overproduction of signal peptidase I in *E. coli*, which will result in a higher level of efficiency of protein translocation and maturation of hybrid secretory proteins. Van Dijl, J.M., *et al.* (1991) *Mol. Gen. Genet. 227*:40-48.

BNSDOCID: <WO 9821346A1 >

There appears to be a limit to which protein translocation can be enhanced in bacteria with current approaches. Most, if not all of the approaches are aimed at elevating or manipulating endogenous bacterial proteins involved in the secretory process. The alteration of bacterial components usually poses a problem because the bacterium itself can monitor and adjust the engineered aberrations, *e.g.*, the secA component. The adjustment or downregulation usually reduces the magnitude of the improvement or abolishes the effect altogether (Oliver (1993) *Mol. Micro.* 7:159; Oliver *et al.* (1990) *J. Bioenerg. Biomembr.* *22*:311).

One approach to overcome this problem is to further enhance transport by elevating the efficiency of protein translocation and/or by creating additional pathways for translocation via the introduction and expression of a central component of the plant plastid protein import apparatus in bacteria.

In the plastid, Bce44B may operate through interaction with a plastid-specific transit signal. Combined use in a protein transport system (*e.g.*, the bacterial transport system), of Bce44B with proteins containing a plastid-specific transit signal could enhance transport of these proteins via Bce44B due to preferential interaction of Bce44B with this transit signal. Accordingly, this would provide a preferred substrate protein for transport with Bce44B, and Bce44B would selectively transport such proteins over endogenous secreted proteins.

In one embodiment, the vectors of this invention can be introduced into suitable hosts where the gene encoding the desired product can be expressed and the product produced. Preferred vectors for this application are T7 promoter-based plasmids; however, almost any type of expression plasmid would

work, including those which are T3 promoter-based, SP6 promoter-based, *lac* promoter-based, *trp* promoter-based, and even inducible promoter-based. The vector may be formed by methods well known to those skilled in the art to which the invention relates. Such methods are described in greater detail in various publications identified herein, but especially Sambrook *et al.*, *supra*, the contents of which are hereby incorporated by reference into the present disclosure in order to provide complete information concerning the state of the art.

Suitable hosts for these methods include *E. coli*, *Bacillus* sp., yeasts, pseudomonads, and cyanobacteria. Gene expression technology in these hosts is described in *Methods in Enzymology*, Volume 185, Goeddel, D.V., ed. (1990).

The resulting host/vector systems can be employed to manufacture the foreign product. The host cells containing the vectors are grown under suitable conditions permitting production of the foreign product and its transport into the periplasm and/or growth medium for recovery. Expression or membrane incorporation can occur under a wide range of conditions suitable for the growth of most hosts. As shown in Example 17, Bce44B-facilitated transport is not inhibited by low temperatures and very likely performs within the normal temperature range of all organisms; thereby requiring no special environmental conditions for its activity.

There are many advantages to the methods of this invention which provide a supplementary transport Bce44B protein or its functional equivalent in the inner and outer plasma membrane of the host organism (single membrane of a yeast host). First, translocation levels of the desired product will increase, resulting in higher yields of harvested product. This can come

about because the product is not retained inside the cell where it can either be degraded or cause reduced growth of the host or both. Further, it is common for overexpressed foreign proteins to become insoluble, making processing of active proteins difficult or even impossible. For instance, proteins often have to be denatured and slowly renatured from insoluble inclusion bodies. The Bce44B protein is soluble when expressed inside a host bacterium and, if it binds precursor proteins or if fused to a precursor protein, it could enhance solubility of foreign proteins expressed in bacteria. Finally, there is evidence that additional export of the product through the bacterial cellular membrane by means of the Bce44B protein may increase expression of the product inside the cell. That is, synthesis of the product is propelled by its removal from the intracellular pool. Thus, in addition to reducing the costs of harvesting the fermentation product, increased yields can result from application of the methods of this invention.

Bacteria and yeasts are important organisms used for the commercial production of molecular products. Examples of such products include amino acids, enzymes, nucleotides, hormones, vitamins, antibiotics, antibodies and the like. More specifically, acid phosphatases, human growth hormones (Chang C., *et al.* (1987) *Gene* 55:189-196), growth factors (Wong, E.Y., *et al.* (1988) *Gene 68*:193-203), human interleukins (Denefle, P., *et al.*, *Gene 85*:499-510), epidermal growth factors (Morioka-Fujimoto, K., *et al.* (1991) *J. Biol. Chem. 266*:1728-1732) and bovine somatotropin (Klein, B.K., *et al.* (1992) *Protein Engin. 5*:511-517) have been described for such applications. More examples are listed at Wrotnowski, C., *Genetic Engineering News 16(12)*:6 (June 15, 1996). Those of skill in the art will

recognize that the above examples are not limiting and illustrate only a few of the products, manufacture of which can be enhanced by the methods of this invention.

The following examples describe specific aspects of the invention to illustrate the invention and provide a description of the methods used to isolate and modify the Bce44B gene and to identify its function in organisms. The examples should not be construed as limiting the invention in any way.

All citations in this application to materials and methods are hereby incorporated by reference.

Example 1

Subfractionation of chloroplasts and purification of chloroplast envelopes

Intact chloroplasts were purified from pea seedlings, *Pisum sativum* (cv. Improved Laxton's), as described by Bartlett, *et al.* (1982) *Methods in Chloroplast Molecular Biology*, (Edelman, M., Hallick, R.B. and Chua, N.H., eds.) pp. 1081-1091 Elsevier Biomedical Press, Amsterdam, or Cline, *et al.* (1985) *J. Biol. Chem. 260*:3691. The growth conditions were identical to those described previously (Ko and Cashmore (1989) *EMBO J. 8*:3187). Pea seedlings from 200 grams of seeds were grown for 9-11 days in growth chambers set at 21°C under fluorescent lighting with 16:8 h light:dark photoperiod. Pea seedlings were harvested and homogenized in cold grinding buffer (50 mM HEPES-KOH pH 7.6, 0.33 M sorbitol, 0.05% bovine serum albumin, 0.1% ascorbate, 1 mM magnesium chloride, 1 mM manganese chloride, 2 mM Na_2EDTA) for 2-3 brief blendings of 5-10 sec at a setting of 5-6 on the Polytron Homogenizer. Percentages for all solutions are wt/vol unless otherwise indicated. All steps were conducted on ice with

chilled equipment and solutions (4°C). The homogenate was then filtered through three layers of Miracloth and the crude chloroplasts collected by centrifugation at 2,800 x g for 3 minutes at 4°C. The crude chloroplast pellet was resuspended in 4 ml of grinding buffer and layered onto a 10-80% Percoll gradient (50 mM HEPES-KOH pH 7.6, 0.33 M sorbitol, 0.05% bovine serum albumin, 0.1% ascorbate, 0.15% polyethylene glycol, 0.05% Ficoll, 0.02% glutathione, 1mM magnesium chloride, 1 mM manganese chloride, 2 mM Na_2EDTA and Percoll). The gradients were centrifuged in a swing-out rotor at 10,000 x g for 10 min at 4°C. The intact chloroplast band near the bottom of the gradient was collected and diluted at least five fold with 1x HS buffer (50 mM HEPES-KOH pH 8.0, 0.33 M sorbitol). The intact plastids were collected by centrifugation at 4,350 x g for 2 minutes. This step was then repeated once more with the pelleted chloroplasts by resuspending in 1x HS. The final pellet was resuspended in 5 ml of 1x HS and an aliquot subjected to chlorophyll analysis. Chlorophyll assays were performed as described by Arnon (1949) *Plant Physiol. 24*:1. Samples were extracted with 80% acetone/20% water (v/v). Insoluble material was removed by centrifugation in a microfuge for 1 minute at high speed. The supernatant was removed for spectrophotometric analysis of chlorophyll according to the Arnon conversion equation. Chloroplast envelopes were subfractionated using the freeze-thaw method and discontinuous or flotation sucrose gradients according to Cline, *et al.* (1985) *J. Biol. Chem. 260*:3691 or Keegstra and Yousif (1986) *Methods Enzymol. 118*:316. The final pellet of intact chloroplasts was collected by centrifugation as before and resuspended in a hyperosmotic solution of 0.6 M sucrose, 10 mM Tricine-NaOH pH 7.5 and 2 mM EDTA. Rupturing of chloroplasts was facilitated by one

cycle of freezing for 1.5 hour at -20°C and thawing at room temperature. A crude envelope fraction was collected by diluting with 2 volumes of 10 mM Tricine-NaOH pH 7.5, 2 mM EDTA, centrifuging at 4,500 x g for 15 minutes to remove the bulk of thylakoids and then recentrifuging the supernatant at 40,000 x g for 30 minutes. The crude envelope pellet was then resuspended in 0.2 M sucrose, 10 mM Tricine-NaOH pH 7.5, 2mM EDTA buffer for discontinuous sucrose gradients. The discontinuous sucrose gradients consisted of 3 ml of 1.0 M sucrose, 3 ml of 0.8 M sucrose and 3 ml of 0.46 M sucrose (all sucrose solutions were made in 10 mM Tricine-NaOH pH 7.5, 2 mM EDTA buffer). The resuspended crude envelope mixture (approximately 2.5 ml in volume per gradient) was then overlaid onto the discontinuous gradient and centrifuged at 180,000 x g for 2 hours at 4°C in a swing-out rotor. The outer and inner envelope fractions were collected from the corresponding interfaces. The outer membrane fraction was located in the 0.46/0.8 M interface and the inner membrane fraction was in the 0.8/1.0 M interface. These fractions were diluted at least five volumes with Tricine-EDTA buffer, and collected by centrifugation as described for the pelleting of crude envelope preparations. Alternatively, the ruptured chloroplasts can be readjusted to 1.3 M sucrose using a 2.6 M sucrose solution if flotation gradients are used. This mixture (approximately 15 ml) was then overlaid with 9 ml of 1.2 M sucrose and 6 ml of 0.3 M sucrose (all made in Tricine-EDTA buffer as above). The resulting gradients were then centrifuged at 113,000 x g for 10-14 hours at 4°C. Total envelopes were collected from the 0.3/1.2 M interface, diluted with Tricine-EDTA buffer and pelleted by centrifugation as described above. All envelope membrane preparations were stored in Tricine-EDTA buffer

DNSDOCID: <WO 9821346A1 | >

at -70°C until use. Total mixed outer and inner chloroplast envelope preparations were used for the immunization protocols. Approximately 50-100 µg of total envelope protein was used for each immunization injection.

Example 2

Antibody preparations

Monospecific type G immunoglobulins used in the various immunological experiments were prepared using female New Zealand white rabbits according to Chua, *et al.* (1982) In: *Methods in Chloroplast Molecular Biology* (Edelman, M., Hallick, R.B., Chua, N.H., eds.) pp.1063-1080. Elsevier Biomedical Press, Amsterdam. The polyclonal antiserum initially used in the identification of cDNA clones encoding Com44/Cim44 (Bce44B is a member of the Com44/Cim44 protein family) was raised against total pea chloroplast envelope proteins as previously described in Ko, K., *et al.* (1992) *J. Biol. Chem. 267*:2986. Antibodies against the COOH terminus of the tomato 44 kDa envelope protein (Tce44) and both NH_2 and COOH termini of the *Brassica napus* 44 kDa envelope protein (Bce44B) were generated as outlined in Wu, *et al.* (1994) *J. Biol. Chem. 269*:32264. These proteins were generated using the T7 RNA polymerase-expressing bacterial overexpression system. The plasmid vector pGEMEX-1 (Promega) and *E. coli* strain JM109(DE3) were used to facilitate overexpression. Protein inclusion bodies were purified according to a procedure in Sambrook, *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Denatured fusion proteins (pGEMEX-1 foreign protein segments fused to T7 gene 10 protein) were prepared and purified by preparative SDS-PAGE prior to extraction and electroconcentration.

Approximately 25 µg of antigens were used for each booster injection. Each injection in addition contained saline solution (150 mM sodium chloride and 10 mM phosphate buffer pH 7.0), 0.1% SDS (w/v) and RIBI adjuvant (RIBI Immunochem Research Inc, Hamilton, Montana). A total volume of 500 µl was injected each time. The initial immunization program occurred over a six-week period and the rabbits were given monthly boosters a week prior to bleeding. Preimmune sera and corresponding IgGs were collected and purified prior to the injection of each rabbit.

Example 3

Identification of cDNA clones encoding 44 kDa proteins

Construction of the *Brassica napus* (cv. Topas) cDNA expression library and the strategy for immunoscreening the *B. napus* cDNA expression library were performed as described in Ko, K., *et al.* (1992) *J. Biol. Chem. 267*:2986 and Ko, K. *et al.*, (1994) *Plant Physiol. 104*:1087. The cDNA expression library was constructed in the phage vector λgt22 (Promega) using polyadenylated RNA isolated from 21 day old *Brassica napus* (cv. Topas) developing seeds. The synthesis of cDNA was facilitated using a kit purchased from Promega. The growth and propagation of the cDNA library was performed according to standard methods such as the ones described in *Molecular Cloning: A Laboratory Manual* (Sambrook, *et al.* (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). The cDNA inserts were retrieved by an *Eco*RI-*Not*I digestion from the selected appropriate recombinant phage DNA and subcloned into pGEM11Z (Promega) for expression in JM109(DE3) and for further analysis.

FINSDOCID: <WO 9821346A1 >

Example 4

Nucleotide sequence analysis

The nucleotide sequence of the Bce44B cDNA insert was determined by the standard dideoxynucleotide chain-termination method using double-stranded DNA templates (*Molecular Cloning: A Laboratory Manual*, Sambrook et al., (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Example 5

Protein analysis

All proteinaceous samples were resolved and analyzed by denaturing SDS-PAGE (Laemmli, (1970) *Nature 227*:680; Towbin, *et al.* (1979) *Proc. Natl. Acad. Sci. USA 76*:4350), electrophoretically transferred onto nitrocellulose filters and immunologically analyzed as described by Hoffman, *et al.* (1987) *Proc. Natl. Acad. Sci. USA 84*:8844. Primary immunoreactions were detected using alkaline phosphatase-conjugated anti-rabbit IgGs (Sigma; Promega). The resulting immunoblots were analyzed and normalized by laser densitometry using the LKB ULTRASCAN XL laser densitometer and the software GELSCAN.

Example 6

Construction of Bce44B expression plasmid

All DNA cloning and expression plasmids discussed were propagated in the *Escherichia coli* strains HB101 or the JM101-109 strain series. The transformation of various bacterial strains was carried out using standard protocols (such as the ones described in *Molecular Cloning: A Laboratory Manual*, Sambrook, *et al.* (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). Plasmid DNAs were isolated from the bacterial strains

harboring the corresponding plasmids using standard protocols (such as the ones described in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, 1989). Gene constructs that express the Bce44B protein can be created, inserted and propagated in a variety of noncommercial or commercially-available plasmids such as the pBLUESCRIPT series (Stratagene), the pBS series (Stratagene), the pGEM and pSP series (Promega) and pT7/T3 series (Pharmacia) if the T7 RNA polymerase bacterial expression system is used to synthesize Bce44B protein. The T7 RNA polymerase gene in the appropriate bacterial strains such as JM109(DE3) or BL21(DE3) is under the control of the IPTG-inducible *lac* promoter. The currently used promoter for expression in the T7 RNA polymerase containing/expressing bacteria is the T7 promoter. Termination sequences such as the T7 terminator can be used in addition to any functionally equivalent sequences present in the gene itself. Other expression systems such as the IPTG-inducible system based on the *lac* promoter can also be used to express the Bce44B protein, for example, the pKK233 series (Clontech) or the pPROK series (Clontech). Any other bacterial expression system that causes the expression of the Bce44B protein in a desirable manner including constitutive expression can also be used. Plasmids usually contain multiple cloning regions for cloning manipulations, an origin of replication and a selectable gene marker such as antibiotic resistance. Expression plasmids additionally contain an appropriate promoter.

All restriction endonuclease digestions were carried out in accordance with the buffers and protocols provided by the manufacturer of each particular enzyme. Restriction enzyme was added to give 5-10 units per microgram of DNA and the reaction mixture was adjusted to the appropriate final volume with water.

The final volumes were usually 20-100 µl and contained 2-10 µg of plasmid DNA. Digestions were thoroughly mixed and carried out for 1 hour at the appropriate suggested temperature. Digested DNA molecules were re-purified by phenol and chloroform:iso-amyl alcohol extraction, centrifugation (usually in a microfuge) and the aqueous layer containing the digested DNA concentrated by precipitation in two volumes of 100% ethanol in the presence of 0.3 M sodium acetate, pH 7.0 or 0.1 M sodium chloride. The phenol used was saturated with 0.1 M Tris-HCl pH 8.0 plus 0.1% (w/v) hydroxquinoline prior to use. The chloroform:iso-amyl alcohol consisted of 24 volumes of chloroform and 1 volume of iso-amyl alcohol. Equal volumes of phenol or chloroform:iso-amyl alcohol were used in each of the organic solvent extraction steps. The DNA precipitates were collected by centrifugation, washed once with 70% ethanol (70% ethanol, 30% water), dried and redissolved in an appropriate volume of water prior to further manipulations.

The Bce44B expression plasmid was made by first retrieving the 1193 base pair cDNA insert from the recombinant phage DNA described above. Retrieval was achieved by *Eco*RI and *Not*I restriction enzyme digestions. The purification of the cDNA insert was carried out using the standard low melting agarose gel and phenol extraction method (*Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). The low melting agarose was supplied by GIBCO-BRL, Gaithersburg, MD, U.S.A. DNA was recovered from appropriate low melting agarose slices by heating at 65°C followed by extraction with phenol that had been prewarmed at 37°C, and centrifugation. The phenol extraction was repeated. The aqueous layer containing the DNA was then adjusted to 0.1 M

sodium chloride and centrifuged for 10 min in a microfuge. The supernatant was then given a chloroform:iso-amyl alcohol extraction followed by precipitation in ethanol as described above. The DNA pellet was then collected by centrifugation, washed with 70% ethanol, dried and resuspended in water. The pGEM11Z plasmid was digested with the same enzymes and dephosphorylated. Phosphatase reactions were carried out by adjusting the restriction digestion reactions with 3.5 µl 1M Tris-HCl, pH 8.0 (per 100 µl reaction) and adding 0.5 unit of calf intestinal alkaline phosphatase. Incubation proceeded for 30 minutes at 37°C and the DNA was then repurified by organic solvent extraction followed by ethanol precipitation as above.

The pOee1-Dhfr plasmid encoding a chimeric chloroplast protein presursor (Wu, C. and Ko. K (1993) *J. Biol. Chem.* *268*:19384-19391) was constructed in the same manner using the same expression vector as pBce44B except that the gene encoding Oee1-Dhfr replaced the gene encoding Bce44B. Plasmid pOee1-Dhfr was used as the expression control plasmid.

The ligation reactions consisted of the two appropriate target DNA molecules, ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 1mM dithiothreitol, 1 mM ATP) and 1-3 units of enzyme. The ligation reaction was carried out at 15°C using T4 DNA ligase from various suppliers. See, *e.g.*, Sambrook, *et al.*, *supra*.

Example 7

Construction of the truncated Bce44B expression plasmids

Carboxyl-terminal deletions of Bce44B were created by exonuclease III/S1 digestion (Henikoff, (1987) *Meth. Enzymol.* *155*:156. The resulting truncated products of these deletion

constructs (designated C1-C5) lacked 10, 24, 197, 234 and 284 amino acids from the COOH terminus, representing deletions of 3, 7, 60, 72 and 80% of the Bce44B protein, respectively. All of these constructs were subcloned into the *Eco*RI and *Sma*I sites of pGEM4. The DNA fragments generated by deletion possessed a blunt end and an *Eco*RI end.

Amino-terminal deletions of Bce44B were generated by exonuclease III/S1 digestion and joining the digested DNA fragments to the DNA sequence for the first 4 amino acids of the RbcS transit peptide (MASM) from *Pisum sativum.* This cloning strategy is similar to that reported earlier (Wu and Ko, (1993) *J. Biol. Chem. 268*:19384; Ko and Cashmore (1989) *EMBO J. 8*:3187). The resulting translation products of these deletion constructs (designated N1-N4) lacked 42, 82, 140 and 181 amino acids from the NH_2 terminus. The amino acid sequences of the fusion site of N1-N4 are MASMISSLSVPPQ-, MASMISSLSVPPSV-, MASMISSLSRLF- and MASMMYPKMI-, respectively. All of these deletion constructs (N1-N4) were made in pGEM4. The N1 deletion construct was then subcloned into the same plasmid vector as employed for Bce44B and the resulting recombinant plasmid was designated pK117. See Example 18.

The K118 fusion construct was made by joining the DNA sequence for the first 23 amino acids of Bce44B to N1. The DNA fragment encoding the NH_2-terminal 23 amino acids and the 5' untranslated region of *Bce44B* was retrieved from C5 using *Eco*RI and *Hinf*I. The *Eco*RI-*Hinf*I DNA fragment was inserted into pGEM4 via *Eco*RI and *Sma*I, after first converting the *Hinf*I site to a blunt end using the Klenow fragment of *E. coli* DNA polymerase I. This resulting vector was used for the construction of K118. The K118 fusion was completed by joining to an *Asp*718-*Hind*III DNA

fragment retrieved from N1 via the *Bam*HI and *Hind*III sites of the above vector. (The *Asp*718 and *Bam*HI sites were first made blunt ends.) The amino acid sequence of the fusion point is -GLGIVPP-. The fusion construct was subcloned in pGEM4 and then transferred to the same expression plasmid as described above. The final recombinant plasmid was named pK118.

Example 8

Expression of Bce44B in *E. coli*

The expression plasmid pBCE44B was transformed into the *E. coli* strain JM109(DE3) using standard calcium chloride methods. Selected colonies were checked for the presence of the plasmid and its quantity in the cell. The resulting strain of JM109(DE3) containing pBCE44B plasmid was recovered and stored in glycerol stocks at -70°C or -20°C until use. The glycerol stocks were made by taking 750 µl of log phase growing cells and mixing in 150 µl of sterile glycerol.

For each expression experiment, the cells were streaked out on LB-ampicillin plates (25 µg/ml) and incubated at 37°C overnight. Colonies formed on the overnight plate were then used to inoculate a liquid culture of LB and ampicillin. Incubation again proceeded with shaking overnight at 37°C. This overnight liquid culture was then used to inoculate a culture to induce expression of Bce44B. The overnight liquid culture was used as an inoculum at a ratio of 1/100 and the freshly inoculated culture allowed to grow with shaking for 2 hours at 37°C.

Induction of expression was achieved by the addition of 10 µl of isopropyl B-D-thiogalactopyranoside (IPTG) (48 mg/ml stock) two hours after inoculation. Lower but sufficient levels of expression can also be obtained without IPTG induction.

ENSDOCID <WO 9821346A1 | >

Expression was allowed to proceed for another 2-3 hours before the cells were collected by centrifugation.

The pelleted cells were then resuspended in SDS-PAGE loading buffer (5% SDS (w/v), 0.1% bromophenol blue (w/v), 20% glycerol (v/v), 1.2 M β-mercaptoethanol, 0.1 M Tris-HCl pH 6.8), boiled for 3 minutes, and centrifuged for 5 minutes in a microfuge before loading onto an SDS-polyacrylamide gel for analysis. Protein gels were visualized by Coomassie Blue staining or more specific protein bands were analyzed by immunoblotting with various specific antibodies as described above.

Example 9

E. coli inner and outer envelope membrane subfractionation

Subfractionation of bacterial membranes was carried out according to Cabelli, *et al.* (1991) *J. Biol. Chem.* *266*:24420. Cells were grown in LB broth containing 25 μg/ml of ampicillin to $O.D._{600}$ = 0.8 - 1.0, pelleted by centrifugation in a JA-20 rotor (Beckman) at 7,000 r.p.m. at 4°C, and resuspended in cold 20% sucrose (w/v) in 10 mM

Tris-HCl pH 8.0, 150 μg/ml DNase I + 150 μg/ml RNase. Cells were then broken twice at a setting of 15,000 p.s.i. in a French Pressure cell. Broken cells were then chilled on ice. Cell debris was removed by centrifugation at 3,000 r.p.m. for 10 min at 4°C in a Beckman JA-20 rotor. The resulting supernatant was subjected to the two-stage separation of the membrane fractions.

Whole membranes were first collected by a low density sucrose gradient consisting of 5 ml 70% sucrose, 12 ml 18% sucrose overlaid with 20 ml of broken cell supernatant in a polyallomer tube. The samples were centrifuged at 23,000 r.p.m. in a SW27 or SW28 rotor (Beckman) for 2 hours at 4°C. Outer and

inner membranes banded at the junction between the 18% and 70% steps and were carefully collected using a glass pipette. The collected whole membrane fraction was subjected to further separation into OM1, OM2, M, and IM by using a 4-step gradient consisting of 3 ml 70% sucrose, 9 ml 64% sucrose, 9 ml 58% sucrose, 9 ml 52% sucrose, and 7 ml of sample. The whole membrane fraction was layered on the top, centrifuged for 4 hours to overnight at 23,000 r.p.m. in a SW27 or SW28 rotor (Beckman) at 4°C. Four bands appeared upon completion of centrifugation; the upper two were reddish, and the lower two were whitish. The lower bands were outer membranes (OM). Bands were carefully collected as separate fractions using a glass pipette. Each collected band was placed in 70Ti polycarbonate tube and diluted with at least 2 volumes of distilled water (to dilute the sucrose concentration to below 20%), then centrifuged at 47,000 r.p.m. at 4°C for 60 minutes in a Beckman Ti 70 rotor. The pellets were resuspended in 50 mM Tris-HCl (pH 7.0) and stored at -80°C until analysis.

The purity of fractions of the inner and outer membrane was determined by detection of cytochrome b1 content, which is mainly in the inner membrane fraction, and of KDO (2-keto-3-deoxyoctonate) content of the peptidoglycan, which is mainly in the outer membrane fraction, using the methods of Deeb and Hager (1969) *J. Biol. Chem. 239*:1024 and Braun and Rehn (1969) *Eur. J. Biochem. 10*:426, respectively. Outer and inner membrane fractions were also analyzed by SDS-polyacrylamide gel electrophoresis and silver staining, confirming the identity of the membrane fractions by visualizing the distinct protein profiles of the two different fractions.

Example 10

Preparation of RSO membrane vesicles

Right-side-out (RSO) vesicles were obtained by the method described in Kim, *et al.* (1994) *Cell* 78:845. Cells were harvested by sedimentation and washed twice in 30 mM Tris-HCl (pH 7.5), 20% sucrose, 1 mM phenylmethylsulfonyl fluoride (PMSF). EDTA (pH 8.0) and lysozyme were added to final concentrations of 10 mM and 0.5 mg/ml, respectively. The cells were then incubated for 40 minutes at 4°C. All subsequent manipulations were performed at 4°C. Spheroplasts were sedimented at 16,000 x g for 15 minutes, and pellets were resuspended in the smallest possible volume (0.5 - 1.0 ml) of 0.1 M Tris - HCl (pH 7.5), 20% sucrose, 20 mM magnesium sulfate, 1 mM PMSF containing DNase and RNase at concentrations of 3-5 µg/ml, using a Teflon and glass homogenizer. The suspension was poured into 300 - 500 vol. of 50 mM Tris - HCl (pH 7.5), 1 mM PMSF and incubated for 15 minutes, when EDTA (pH 8.0) was then added to 10 mM, and incubation was continued for an additional 15 minutes. $MgSO_4$ was added to 15 mM, and incubation was continued for an additional 15 minutes. The lysate was sedimented at 16,000 x g for 30 min, and the pellet was resuspended by vigorous homogenization in a solution of 50 mM Tris - HCl (pH 7.5), 10 mM EDTA, 1 mM PMSF. This preparation was sedimented at 45,000 x g for 20 minutes, and the pellet was resuspended by homogenization in 50 mM Tris - HCl (pH 7.5), 10 mM EDTA, 1 mM PMSF. The sample was sedimented at 800 x g for 30 min, and the yellowish, milky supernatant fluid was carefully decanted and sedimented at 45,000 x g for 20 minutes. Low speed sedimentation was repeated on the 800 x g pellet fraction after the extensive homogenization in order to obtain a maximal yield of RSO membrane vesicles. The resulting RSO

samples were resuspended into aliquots, and stored at -80°C until use.

Example 11

Preparation of ISO membrane vesicles

Inside-out (ISO) vesicles were prepared according to Kim and Oliver (1994) *FEBS Lett.* 339:175. Cells were harvested by sedimentation, washed once in 50 mM Tris-HCl (pH 7.5), 5 mM $MgSO_4$ at 6 ml/g wet weight of cells, and DNase I was added to 10 µg/ml. The cell suspension was passed through the French Pressure cell twice at 5,000 p.s.i. Unbroken cells were removed by sedimentation at 10,000 x g for 10 minutes, and a membrane pellet was obtained after sedimentation at 12,000 x g for 2 hours. Membranes were resuspended in 50 mM Tris - HCl (pH 7.5), 1 mM dithiothreitol, 10% glycerol, and 250 mM sucrose and stored at -80°C until use.

Example 12

Thermolysin digestion of ISO and RSO membrane vesicles

Both RSO and ISO membrane vesicles were adjusted to 3 mg protein/ml in the washing buffers: 50 mM Tris-HCl (pH 7.5) for RSO membrane vesicles; 50 mM Tris-HCl (pH 7.5), 1 mM dithiothreitol, 10% glycerol, 250 mM sucrose for ISO membrane vesicles. Reactions were initiated by adjustment to 0.5 mg of thermolysin/ml prepared in 50 mM Tris-HCl (pH 7.5) containing 1 mM $CaCl_2$. The protease treatment scheme was performed as reported by Cline, *et al.* (1984) *Plant Physiol.* 75:675. After incubation on ice for 25 minutes, reactions were terminated by adjusting reaction mixtures to 10 mM EDTA. Control vesicles were subjected to the same treatment with the exception that control

BNSDOCID: <WO 9821346A1 >

vesicles did not receive thermolysin. Triton X-100 was added to a final concentration of 0.1% to one thermolysin digestion group, as another control. An equal volume of 2x SDS loading buffer was then added to the reaction mixtures. Samples were boiled for 3 minutes, separated by 10% SDS-polyacrylamide gel electrophoresis and immunoblotted with specific anti-Bce44B protein antibodies.

Example 13

Induction and transport of alkaline phosphatase in *E. coli* JM109 (DE3)

The induction and analysis of alkaline phosphatase was performed as described by Torriani, *et al.* (1960) *Biochim. Biophys. Acta. 38*:460. Bacterial cells were grown in LB-ampicillin (25 μg/ml) broth till $O.D._{600}$ = 0.8 - 1.0, diluted 50x into M9 minimal broth (0.06 M K_2HPO_4, 0.03 M KH_2PO_4, 7.5 mM $(NH_4)_2SO_4$, 2 mM sodium citrate.$2H_2O$, pH 7.0) and in M9 (-P_i) broth (phosphate starvation broth) which was identical to the M9 minimal broth except that it contained 0.15 M Tris instead of phosphates. Both media contained 0.1% glucose, 0.2% glycerol, and 1 mM twenty amino acid mixture as additional components. A low concentration of phosphate (50 μM) was present in the M9 (-P_i) minimal broth to provide enough essential phosphorus for cell growth. Cells were collected at 10 hour and 15 hour growing periods and adjusted to the same cell number by using the $O.D._{600}$ values. Samples were then processed in 1x SDS loading buffer, subjected to 10% SDS-polyacrylamide gel electrophoresis and immunoblotted with anti-alkaline phosphatase antibodies to visualize the alkaline phosphatase profiles in both Bce44B-expressing and Oee1-Dhfr-expressing cells.

The results showed that Bce44B-expressing cells grow more slowly than the Oee1-Dhfr-expressing cells even under normal phosphate levels. In the M9 minimal media with 50 μM phosphate (phosphate starvation levels), the Bce44B-expressing cells gave rise to two peaks (10 and 15 hours) during the growing period. Immunoblots were made comparing the amount of processed alkaline phosphatase from the two types of cells at the time of the two growth peaks both with and without phosphate starvation.

Example 14

Analysis of changes in the transport of β-lactamase

Changes to β-lactamase transport were monitored by two different approaches: 1) the ability to form colonies on solid LB-agar plates containing increasing concentrations of ampicillin (50 μg/ml to 3.0 mg/ml); and 2) the level of transported and processed β-lactamase detected by immunoblotting total protein samples of cells grown in LB-broth containing increasing ampicillin concentrations.

Bce44B- and Oee1-Dhfr-expressing cells were grown in LB broth containing 25 μg/ml ampicillin to OD_{600}=0.8-1.0. Cultures were collected and OD_{600} values measured to adjust by dilution the cell numbers (usually 500-700 cells) to approximately the same numbers before plating onto LB agar plates containing increasing amounts of ampicillin. Plates were incubated at appropriate temperatures overnight. Colony numbers on each plate with increasing ampicillin concentrations were compiled and compared to assess the level of ampicillin resistance based on the ability to form colonies.

INSDOCID: <WO 9821346A1>

The same cells were also used to determine the level of transported β-lactamase grown in media containing increasing concentrations of ampicillin. Cultures were diluted 50X into LB broth with increasing amounts of ampicillin (50 μg/ml to 3.0 mg/ml) and allowed to grow up to 4 hours. The level of transported β-lactamase was monitored by immunoblotting analysis of samples taken at time intervals of 0 hr, 0.5 hr, 1.0 hr, 1.5 hr, 2.0 hr, 3 hr and 4 hr. Immunoblot analysis was performed on total cellular protein samples with anti-β-lactamase antibodies as outlined above.

Example 15

Sodium azide sensitivity analysis

Sensitivity of the bacterial strains to sodium azide was determined in the same manner as outlined in Oliver, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8227. Changes to β-lactamase transport activity in response to sodium azide was used as a monitor of differences in the protein translocation activity of Bce44B- versus Oee1-Ohfr-expressing cells. Both Bce44B-expressing and Oee1-Dhfr-expressing cells were grown in LB broth with an ampicillin concentration of 25 μg/ml till $O.D._{600}$ = 0.8 - 1.0, diluted 50x into the LB broth containing increasing concentrations of sodium azide (0 to 1.0 mM) and incubated for 2.0 hours. Samples were then adjusted to the same cell numbers by measuring $O.D._{600}$ values at the time of collection, subjected to 10% SDS-PAGE and immunoblotted with anti-β-lactamase antibodies (5 PRIME 3 PRIME, Inc., Boulder, CO).

The same strains were also assessed using 0.5 mM sodium azide over a period of 2.0 hours. The time course study was

carried out in the same manner as above where the cells were inoculated in LB-ampicillin broth containing 0.5 mM sodum azide. Samples were collected and monitored by immunoblotting at time intervals of 0, 0.5, 1.0, 1.5 and 2.0 hr as outlined above. Control experiments were conducted without sodium azide.

Example 16

Preparation of spheroplasts and chemical cross-linking studies

The chemical cross-linking of bacterial proteins was performed as described by Akita, *et al.* (1990) *J. Biol. Chem.* *265*:8164. The Bce44B-expressing cells and Oee1-Dhfr-containing cells were grown to $O.D._{600}$ = 0.8 - 1.0, harvested by sedimentation at 3,000 x g for 5 minutes, washed twice in 30 mM Tris-HCl (pH 7.5), and resedimented by centrifugation as above. The cells were resuspended in ice-cold 30 mM Tris-HCl (pH 7.5), 20% sucrose, 1 mM PMSF, 10 mM EDTA (pH 8.0), and 0.5 mg/ml lysozyme, followed by incubation for 40 minutes at 4°C. The resulting spheroplasts were sedimented at 16,000 x g for 15 minutes, washed twice with 20% sucrose, 20 mM HEPES-NaOH (pH 7.7) and resuspended in the same buffer. The chemical crosslinker, EDAC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide methiodide), was then added to a concentration of 12.5 mM. The cross-linking reaction was carried out at 25°C for 50 minutes, and quenched for 10 minutes by the addition of 200 mM Tris-HCl (pH 7.5). The cross-linked complexes were resolved by 10% SDS-polyacrylamide gel electrophoresis and immunoblotted separately with specific antibodies against the tetracycline efflux protein, Tet, Bce44B and β-lactamase.

Immunoblots of the cross-linked complexes of the Bce44B-expressing cells showed that a high molecular weight cross-

linking-generated band of approximately 70 kDa immunoreacted with antibodies against Bce44B and β-lactamase.

Example 17

Complementation analysis of Bce44B and secA mutant bacterial strain

The pBCE44B and pOEE1DHFR plasmids were introduced into *E. coli* strain MM52 (Oliver and Beckwith, (1981) *Cell 25*:765) by electroporation using the BioRad Gene Pulser and the supplied protocols for *E. coli* strains. Electrocompetent cells were prepared according to the supplied protocol. The MM52 strain is a derivative of MC4100 with a characterized temperature-sensitive mutation in the *secA* gene, which encodes a component of the bacterial protein secretory machinery. The mutation is termed *secATs51*. The plasmid-harboring MM52 strains were grown in LB-ampicillin broth and analyzed in the same manner as the JM109(DE3) strains described above.

Example 18

Characterization of bacterial strains harboring truncated Bce44B

The pK117 and pK118 plasmids, Figures 4 and 5, respectively, were introduced into *E. coli* strain JM109(DE3) and analyzed using the same criteria as described above for the Bce44B-expressing strains.

Example 19

Construction of a CaMV-Bce44B-NOS Transgene for Expression in Plants

Transgenic plants (*Brassica napus, Arabidopsis thaliana* and *Lycopersicon esculentum*) with altered plastid protein transport

capacity were produced by introducing a transgene construct consisting of DNA encoding the Bce44B protein and a 5' untranslated region immediately upstream of the Bce44B coding sequence. See Figure 6. Expression of the construct was obtained by linking the cauliflower mosaic virus 35S promoter (Odell, J.T., *et al.* (1985) *Nature 313*:810) upstream. Transcriptional termination was facilitated by the gene encoding nopaline synthase (NOS) found in Agrobacterium Ti plasmids or their derivatives such as pBI101 (Clontech). Construction of the 35S CAMV-Bce44B-NOS transgene construct is diagrammed in Figure 6.

The transgene cloning process was initiated by the construction of a vector containing DNA encoding the 35S CAMV promoter. The DNA cloning procedures generally used have been described in the examples *supra*. The 35S CAMV constitutive promoter was retrieved as an *Eco*RI-*Hind*III DNA fragment (approximately 450 base pairs in length) from pCAMV (A.R. Cashmore, Univ. Pennsylvania, Philadelphia, PA). The *Hind*III restriction site was converted to a blunt end with Klenow fragment and the DNA fragment inserted into the *Eco*RI and *Sma*I sites of pGEM4 (Promega). This vector, designated pCAMV2, was used to fuse the 35S CAMV promoter to the Bce44B coding region by retrieving Bce44B as a 1,193 base pair *Eco*RI-*Hind*III DNA fragment from pBce44B (Figure 3) and inserting it into the *Xba*I-*Hind*III sites of pCAMV2. The *Eco*RI and *Xba*I sites were made into blunt end sites using Klenow fragment and the resulting vector was designated pCAMV-Bce44B (Figure 6). The 3' transcription termination sequence from the NOS gene was then added to the linked pCAMV and Bce44B sequences by inserting an *Eco*RI-*Hind*III DNA fragment (approximately 260 base pairs in length) into the

*Not*I-*Hind*III sites of pCAMV-Bce44B, following which the *Eco*RI and *Not*I sites were made blunt using Klenow fragment. The NOS transcription termination sequence was retrieved from the pBI121 binary vector (Clontech) as a *Sac*I-*Eco*RI DNA fragment (approximately 260 base pairs in length). To obtain the appropriate restriction sites for insertion of the NOS transcription termination sequence into pCAMV-Bce44B, the *Sac*I-*Eco*RI DNA fragment was first inserted into the *Sac*I-*Sma*I sites of pGEM4 (this vector was designated pNOS for reference purposes) and then retrieved back out as the *Eco*RI-*Hind*III DNA fragment described above. The resulting transgene-containing vector was designated pCAMV-Bce44B-NOS (Figure 6). The 35SCAMV-Bce44B-NOS transgene was then transferred as a *Bam*HI DNA fragment from pCAMV-Bce44B-NOS to the *Bam*HI site of the binary vector pEND4K (Klee, H., *et al.* (1985) *Biotechnology* 3:637; Horsch, R.B., *et al.* (1985) *Science 227*:181; Holsters, M., *et al.* (1987) *Mol. Gen. Genet. 163*:181). The transgene-containing *Agrobacterium* binary vector was designated pEND4K-CAMV-Bce44B-NOS (Figure 6).

All ligation steps were carried out at 15°C overnight using T4 DNA ligase from various suppliers. All steps of the gene construction process were carried out using the standard $CaCl_2$ bacterial transformation protocol and the *E. coli* host strain HB101. The recombinant plasmids were propagated in HB101 and isolated using standard techniques. Resolution of DNA fragments was facilitated by using standard agarose and polyacrylamide gel electrophoresis techniques. See, Sambrook, *et al.*, *supra*.

BNSDOCID: <WO 9821346A1 >

Example 20

Transformation of Plants

The pEND4K-CAMV-Bce44B-NOS vector was introduced into *Agrobacterium tumafaciens* by the freeze-thaw method. Competent *Agrobacterium* strains (such as LBA4404 or GV3101) were obtained by inoculating 50 ml of LB broth containing the appropriate antibiotics (50μg/ml rifampicin for LBA4404 or 100 µg/ml gentamycin and 150 µg/ml rifampicin for GV3101) with 500 μl of an overnight culture, incubating at 28°C with vigorous shaking until the optical density at 650 nm reached 0.7. Cells were harvested by centrifugation at 2000 x g for 5min at 4°C, washed in ice cold 0.1M $CaCl_2$ and finally resuspended in 1 ml of ice cold 20 mM $CaCl_2$. A 150 μl aliquot of competent LBA4404 or GV3101 cells was removed, mixed with 1 μg of plasmid DNA in a microfuge tube, and immediately frozen in liquid nitrogen. The cells were incubated at 37°C in a water bath or thermostat block for 5 min, 1 ml of LB broth was added, and the mixture incubated at 28°C with shaking for 3h. Cells were recovered by centrifugation at 2000xg for 5 min and resuspended in 100 μL of LB before plating on LB plates containing appropriate levels of antibiotics for selection of kanamycin-resistant cells (100 μg/ml kanamycin and 50 μg/ml rifampicin for the LBA4404 strain or 50 μg/ml kanamycin, 150 μg/ml rifampicin and 100 μg/ml gentamycin for the GV3101 strain). After a 2-4 day incubation at 28°C, kanamycin-resistant colonies growing on the plates were selected and the presence of the pEND4K-CAMV-Bce44B-NOS plasmid confirmed in plasmid preparations prepared from single colonies as described below.

Three ml of LB broth containing appropriate levels of antibiotics as described above for the selection process were inoculated with a single kanamycin-resistant colony and incubated from overnight to 2 days at 28°C with shaking. A 1.5 ml sample of this culture was placed in a microfuge tube and centrifuged for 30 sec to pellet the cells. The pellet was resuspended in 0.1 ml of GTE solution (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM Na_2EDTA) and 4 mg/ml lysozyme, then incubated at room temperature for 10 min. Phenol (30 μl), previously equilibrated with 2 vols of 1% (w/v) SDS, 0.2N NaOH, was added. The mixture was vortexed gently until viscous and incubated at room temperature for 10 min. The lysed cells were neutralized with 3M sodium acetate pH 4.8 (150 μl) and incubated at -20°C for 15 min. The mixture was centrifuged for 3 minutes in a microfuge and the supernatant transferred to a fresh microfuge tube. Two volumes of ethanol were added and the mixture incubated at -80°C for 15 minutes, prior to centrifugation for 3 minutes, after which the DNA pellet was resuspended in 90 μl of water. Ten μl of 3M sodium acetate, pH 7.0, were added, then an equal volume of phenol/chloroform, before vortexing the mixture. After a 5 min centrifugation in a microfuge, the supernatant was transferred to a fresh tube and the DNA precipitated by adding 2 volumes of 100% ethanol. Following a 10 min centrifugation, the pellet was washed with 70% ethanol, dried and resuspended in 50 μl of TE (10 mM Tris-HCl pH 8.0, 1 mM Na_2EDTA).

Restriction endonuclease digestion analysis and Southern blot analysis of the pEND4K-CAMV-Bce44B-NOS plasmid preparation was carried out in accordance with procedures described in Molecular Cloning: A Laboratory Manual (Sambrook, *et al.*, *supra*).

Agrobacterium strains containing the pEND4K-CAMV-Bce44B-NOS plasmid were used to transform plants following a leaf disc transformation protocol, such as the one described by Horsch, *et al.* (1985) *Science 227*:1229), or by *in planta* methods which can include the seed transformation protocol reported by Feldmann and Marks (1987) *Mol. Gen. Genet. 208*:1-9, the more recently updated seed transformation protocol described by Feldmann (1992) *In: Methods in Arabidopsis Research*, (eds. C. Koncz, N.H. Chua, J. Schell), pp 274-289) and Chee, P.P. (1994, US Patent Serial No. 5,169,770), the vacuum infiltration approach reported by Bechtold, N., *et al.* (1993) *CR Acad. Sci. Paris/Life Science 316*:118) and the approach employing the inoculation of wound sites in the primary and secondary influorescence shoots (Katavic, V., *et al.* (1994) *Mol. Gen. Genet. 245*:363). Plants, such as *Lycopersicon esculentum* and *Brassica napus*, used for leaf disc transformation were grown in a greenhouse maintained at 18-24°C and supplemented with fluorescent and incandescent lights. Young leaves were excised and surface sterilized in 10% (v/v) sodium hypochlorite, 0.1% (v/v) Tween, then rinsed 4 times with sterile, deionized water. From this point on, standard aseptic techniques for the manipulation of the sterile material and media were used. Leaf discs, 6 mm in diameter, were made with the aid of a sterile paper punch and incubated for 10-20 min in a 1:5 dilution of cultures of *Agrobacterium* harboring the pEND4K-CAMV-Bce44B-NOS construct. Excess bacteria were removed from the leaf discs by briefly blotting on sterile filter paper before the discs were transferred to petri dishes containing "shoot medium" (Horsch, *et al.* (1988) In: *Plant Molecular Biology* (Eds. S.B. Gelvin, R.A. Schilperoot) Kluwer Acad. Publishers, A5:1-9). Petri plates were sealed with parafilm and incubated in

a growth chamber at 24°C equipped with "grow" mixed fluorescent tubes. After 2 days, *Agrobacterium* growing on the discs were killed by washing in 500 mg/ml Cefotaxime in liquid "shoot medium" and the discs transferred to fresh "shoot medium" containing 500 mg/ml Cefotaxime and 100 mg/l kanamycin to select for growth of transformed plant cells.

Leaf discs were incubated for 3-5 weeks under the same conditions and transferred to fresh medium on a weekly basis. Green shoots emerging from the leaf discs were excised and transferred to "root medium" containing appropriate plant species-specific levels of kanamycin (Horsch, *et al.* (1988), *supra*). Shoots which rooted in the presence of kanamycin were selected for further propagation and for further verification via a variety of approaches including the assessment of NptII activity (McDonnell, R.E., *et al.* (1987) *Plant Mol. Biol. Rep.* 5:380), polymerase chain reaction techniques to detect transgene incorporation, and RNA expression. Sterile transformants were transferred to soil for propagation, selfed, and seeds collected for further examination.

Transgenic plants can also be generated via *in planta* methods. *In planta* methods which circumvent tissue culture requirements can be more cost effective and avoid potential mutations associated with plant tissue culture.

For the seed transformation protocol (Feldmann, K., (1992) *In: Methods in Arabidopsis Research* (eds. C. Koncz, N.H. Chua, J. Schell) pp 274-289), 2.4 grams of *Arabidopsis thaliana* (ecotype Wassilewskija) seeds were surface sterilized for 8 minutes in 5.25% (w/v) sodium hypochlorite containing 0.15% (v/v) Tween 20, then rinsed 6-7 times with sterile water. The seeds were apportioned among 40 125 ml flasks containing 40 ml cocultivation

medium (CCM) (1X MS salts, 40 g/l sucrose, 10 mg/l thiamine, 0.5 mg/ml pyridoxine, 0.5 mg/ml nicotinic acid and 100 mg/l inositol, pH 6.0 with KOH). These seed-containing flasks were shaken (190 rpm, 22°C) and the seeds allowed to imbibe with constant lighting (approximately 250 lux) for a period of 10-18 hours. The flasks were then shaken vigorously to loosen clumped seeds and seeds stuck to the walls of the flasks. A 5 ml aliquot of a fresh log phase culture (optical density of 0.75) of the appropriate *Agrobacterium* strain was added to each seed-containing flask. The *Agrobacterium* culture was grown without selective antibiotics at 28°C in LB broth. The flasks were shaken for a 24 hour cocultivation period until the *Agrobacterium* reached stationary phase. Seeds were collected by filtration on filter paper, allowed to dry for 30 minutes in a fume hood, and sown within one hour of drying the seeds. Seeds scraped off the filter paper were sown onto presoaked soil mixtures and covered with plastic wrap containing aeration slits. The seeds and resulting plants were maintained at 21°C in growth chambers with a 16:8 light dark photoperiod (illumination at 100 μmoles per meter^2 per second) until maturity. The plants were watered daily with a nutrient solution such as Hoaglund's medium (per liter: $Ca(NO_3)_2 \cdot H_2O$, 1.1g; $MgSO_4 \cdot H_2O$, 738 mg; KNO_3, 505 mg; NH_4NO_3, 120 mg; KH_2PO_4, 8 mg; Chelated iron (supplied as Sequestrine 138-Fe, Ciba-Geigy Corp.), 15 mg; H_3BO_3, 2.1 mg; $MnCl_2 \cdot H_2O$, 1.4 mg; $ZnSO_4 \cdot 7H_2O$, 165 μg; $H_3MoO_4 \cdot H_2O$, 68 μg; $CuSO_4 \cdot 5H_2O$, 60 μg, adjusted to pH of 5.7-6.0 in soil with H_2SO_4). After approximately seven weeks, the plants were allowed to dry and the T2 seeds were bulk harvested. These seeds were then used for further screening and analysis.

Phenotypic observations of the primary generation of transgenic plants (*Lycopersicon esculentum*, *Brassica napus* and

Arabidopsis thaliana) indicate that the transgenic plants produce seeds approximately twice the size of their wildtype counterparts or the control plants transformed with *Agrobacterium* harboring only the pEND4K plasmid. This phenotypic change may be a protein transport effect caused by the altered levels of Bce44B in the envelopes of plastids throughout all parts of the plant or at least in the seed tissues. Similar results have observed with transformed *Brassica* seeds.

Example 21

Monoclonal antibodies

Monoclonal antibodies (Mab's) specific for isolated or recombinant Bce44B proteins or polypeptides are useful, for example, for diagnostic purposes such as for determining Bce44B protein levels in the identification of cells which express Bce44B. To produce these antibodies, purified Bce44B protein or polypeptide is prepared. The protein or polypeptide is produced in bacterial cells as a fusion protein with glutathione-S-transferase using the vector pGEX2 (Pharmacia). This permits purification of the fusion protein by GSH affinity chromatography. In another approach, Bce44B protein is expressed as a fusion protein with the bacterial maltose binding domain. The fusion protein is thus recovered from bacterial extracts by passing the extract over an amylose resin column followed by elution of the fusion protein with maltose. For this fusion construct, the vector pMalC2, commercially available from New England Biolabs, is used. This vector has been used in the past, for example, to overexpress nuclear receptor proteins which were recovered in high yields for functional studies and the production of receptor specific antisera (Ohno, *et al.* (1993)

TACTCATCTG GGTGGAGAAG ACTTTGACCA GCGTGTCATG GAACACTTCA TCAAACTGTA 2580

CAAAAAGAAG ACGGGCAAAG ATGTCAGGAA GGACAATAGA GCTGTGCAGA AACTCCGGCG 2640

CGAGGTAGAA AAGGCCAAGG CCCTGTCTTC TCAGCATCAA GCAAGAATTG AAATTGAGTC 2700

CTTCTATGAA GGAGAAGACT TTTCTGAGAC CCTGACTCGG GCCAAATTTG AAGAGCTCAA 2760

CATGGTATGT TCCTTGTTTT CTGCTTTGCT AATGAGATCT CCTTAGACTC TGAATTCAGG 2820

ACATTGCATC TAGATACTTA GATAACAGAC ATCACAGTAA CCATGTCTTT TTTCTAGGAT 2880

CTGTTCCGGT CTACTATGAA GCCCGTCCAG AAAGTGTTGG AAGATTCTGA TTTGAAGAAG 2940

TCTGATATTG ATGAAATTGT TCTTGTTGGT GGCTCGACTC GAATTCCAAA GATTCAGCAA 3000

CTGGTTAAAG AGTTCTTCAA TGGCAAGGAA [illegible] AGATGAAGCT 3060

GTAGCGTATG GTGCTGCTGT CCAGGCTGGT [illegible] TACAGGTAGG 3120

TCATCATCGC AGCATCTTTC TTAGTGATTC [illegible] CGGTACCCCT 3180

ATTGCTTTAG AAAATACCAG AATATGAGCA ACAAGGTCAC ACAGCTAGTA AAGGGTATAA 3240

GTGAAGACAA GACTGGGGTA GTCTCCAAGA TCATTAGCAA CTGTTTAATT CACTGCCTTT 3300

AAAATGTGTG TGTTAGAACC TAACCAAATG TTAGAGAGAT AAACTTTACA TAGCTCATAG 3360

GGAGAACTTG AATTAAAAGT TAAATAACTT ATCCTTACAG GTGACCTGGT ACTGCTTCAT 3420

GTATGTCCCC TTACACTTGG TATTGAAACT GTAGGAGGTG TCATGACCAA ACTGATTCCA 3480

AGTAATACAG TGGTGCCTAC CAAGAACTCT CAGATCTTTT CTACAGCTTC TGATAATCAA 3540

CCAACTGTTA CAATCAAGGT CTATGAAGGT AATTACCTTA AGTTTGGTTA ATATCATGGC 3600

TTTTTTTTTG AGATGAAGTC TTGCTCTGTT GCCCAGGCTG GACTGCAGTG GCACGATCTC 3660

GGCTCACTGC AAATTCTGTC TCCCGGGTTC AAGTGATTCT CCTGCCTCAG CCTCCAGAGT 3720

AGCTGGATTA CAGCCTGACC ACCACACCTG GCTAATTTCT GTATTTTTAG TAGAGGATGG 3780

GCTTTCACCA TGTTTCCCAG GCTGGTCTCC AACTCCTGAC CTCAGGTCAT CTGCCTGCCT 3840

CCACCGTCCC GAAAGTACTG GGATTATAGC GTGAGCCACC ACGCCAGATC TATCTATCAT 3900

GGCATATTTT AAAAGAACAT GACTTAATAT GTCCTATTGA AATGGCTAGG GAACTAAGTA 3960

ACTGCTGTTT TCAGATGGAG GTCTTAATTT GAATAATGTT GATATTAGAT ATTTAGCATT 4020

CTTTTTTTTT TTTTTTTAAT.GGAGTCTTGC TCTGTCGCCT AGGCTGGGGT GCAGTGGCAT 4080

BNSDOCID: <WO 9409012A1 >

GACTTGCAAC CTCTGCCTCC CGAATAGCTG GGATTACAGG TGCCCACCAT CACGCCCGGC 4140

TAAGTTTTGT ATTTTTAGTA GAGGCGAGTT TCGCCATGTT GGCCAGGCTG GTCTTGAACC 4200

CCTAACCTCA GTGATCCCAC GGTCACCGAC CTGGCCTCCC AAAAGTACTG TACCCAGCCA 4260

ATGATTAGCA TTCTCACTAA TAATAGCATC TGAGCTGGCT CCTAGAGTAC AAGAAAAAGG 4320

AGTTCACAGT ACTTTAAAAT AGATAAAATT CAGTTGAGTT AGTAACCTAA CTCATTGTTA 4380

GTACTAGTTG CTGCTCCTTG TAGACCAATA TGAAATTACT TTTAGCTCGA TAAAACCAAA 4440

AGTGTCACTT TATGCTTCAG ACTGAAATGC GGGGATCTAG ATGTGCTAAT GCTTGTCAGT 4500

AACAACTAAC AAGTTTTTCT GTATGTAACT TCTAGGTGAA AGACCCCTGA CAAAAGACAA 4560

TCATCTTCTG GGTACATTTG ATCTGACTGG AATTCCTCCT GCTCCTCGTG GGGTCCCACA 4620

GATTGAAGTC ACCTTTGAGA TAGATGTGAA TGGTATTCTT CGAGTGACAG CTGAAGACAA 4680

GGGTACAGGG AACAAAAATA AGATCACAAT CACCAATGAC CAGAATCGCC TGACACCTGA 4740

AGAAATCGAA AGGATGGTTA ATGATGCTGA GAAGTTTGCT GAGGAAGACA AAAAGCTGAA 4800

GGAGCGCATT GATACTAGAA ATGAGTTGGA AAGCTATGCC TATTCTCTAA AGAATCAGAT 4860

TGGAGATAAA GAAAAGCTGG GAGGTAAACT TTCCTCTGAA GATAAGGAGA CCATGGAAAA 4920

AGCTGTAGAA GAAAAGATTG AATGGCTGGA AAGCCACCAA GATGCTGACA TTGAAGACTT 4980

CAAAGCTAAG AAGAAGGAAC TGGAAGAAAT TGTTCAACCA ATTATCAGCA AACTCTATGG 5040

AAGTGCAGGC CCTCCCCCAA CTGGTGAAGA GGATACAGCA GAAAAAGATG AGTTGTAGAC 5100

ACTGATCTGC TAGTGCTGTA ATATTGTAAA TACTGGACTC AGGAACTTTT GTTAGGAAAA 5160

AATTGAAAGA ACTTAAGTCT CGAATGTAAT TGGAATCTTC ACCTCAGAGT GGAGTTGAAA 5220

CTGCTATAGC CTAAGCGGCT GTTTACTGCT TTTCATTAGC AGTTGCTCAC ATGTCTTTGG 5280

GTGGGGGGGA GAAGAAGAAT TGGCCATCTT AAAAAGCGGG TAAAAAACCT GGGTTAGGGT 5340

GTGTGTTCAC CTTCAAAATG TTCTATTTAA CAACTGGGTC ATGTGCATCT GGTGTAGGAG 5400

GTTTTTTCTA CCATAAGTGA CACCAATAAA TGTTTGTTAT TTACACTGGT CTAATGTTTG 5460

TGAGAAGCTT 5470

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 66..2005

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAGGCAGCTG CCGGGCATTA GTGTGGTCTC GTCGTCAGCG CAGCTGGGCC TACACACAAG      60

CAACC ATG TCT AAG GGA CCT GCA GTT GGC ATT GAT CTC GGC ACC ACC          107
      Met Ser Lys Gly Pro Ala Val Gly Ile Asp Leu Gly Thr Thr
        1               5                  10

TAC TCC TGT GTG GGT GTC TTC CAG CAT GGA AAG GTG GAA ATT ATT GCC        155
Tyr Ser Cys Val Gly Val Phe Gln His Gly Lys Val Glu Ile Ile Ala
 15                  20                  25                  30

AAT GAC CAG GGT AAC CGC ACC ACG CCA AGC TAT GTT GCT TTC ACG GAC        203
Asn Asp Gln Gly Asn Arg Thr Thr Pro Ser Tyr Val Ala Phe Thr Asp
                35                  40                  45

ACA GAG AGA TTA ATT GGG GAT GCG GCC AAG AAT CAG GTT GCA ATG AAC        251
Thr Glu Arg Leu Ile Gly Asp Ala Ala Lys Asn Gln Val Ala Met Asn
            50                  55                  60

CCC ACC AAC ACA GTT TTT GAT GCC AAA CGT CTG ATC GGG CGT AGG TTT        299
Pro Thr Asn Thr Val Phe Asp Ala Lys Arg Leu Ile Gly Arg Arg Phe
        65                  70                  75

GAT GAT GCT GTT GTT CAG TCT GAT ATG AAG CAC TGG CCC TTC ATG GTG        347
Asp Asp Ala Val Val Gln Ser Asp Met Lys His Trp Pro Phe Met Val
    80                  85                  90

GTG AAT GAT GCA GGC AGG CCC AAG GTC CAA GTC GAA TAC AAA GGG GAG        395
Val Asn Asp Ala Gly Arg Pro Lys Val Gln Val Glu Tyr Lys Gly Glu
 95                  100                 105                 110

ACA AAA AGT TTC TAC CCA GAG GAA GTG TCC TCC ATG GTT CTG ACA AAG        443
Thr Lys Ser Phe Tyr Pro Glu Glu Val Ser Ser Met Val Leu Thr Lys
                115                 120                 125
```


BNSDOCID: <WO 9408012A1 >

```
ATG AAG GAA ATT GCA GAA GCA TAC CTC GGA AAG ACT GTT ACC AAC GCT      491
Met Lys Glu Ile Ala Glu Ala Tyr Leu Gly Lys Thr Val Thr Asn Ala
            130                 135                 140

GTG GTC ACA GTG CCC GCT TAC TTC AAT GAC TCT CAG CGA CAG GCA ACA      539
Val Val Thr Val Pro Ala Tyr Phe Asn Asp Ser Gln Arg Gln Ala Thr
        145                 150                 155

AAA GAT GCT GGA ACT ATT GCT GGC CTC AAT GTA CTT CGA ATC ATC AAT      587
Lys Asp Ala Gly Thr Ile Ala Gly Leu Asn Val Leu Arg Ile Ile Asn
    160                 165                 170

GAA CCA ACT GCT GCT GCT ATT GCT TAT GGC TTA GAT AAG AAG GTC GGA      635
Glu Pro Thr Ala Ala Ala Ile Ala Tyr Gly Leu Asp Lys Lys Val Gly
175                 180                 185                 190

GCT GAA AGG AAT GTG CTC ATT TTT GAC TTG GGA GGT GGC ACT TTT GAT      683
Ala Glu Arg Asn Val Leu Ile Phe Asp Leu Gly Gly Gly Thr Phe Asp
                195                 200                 205

GTG TCA ATC CTC ACT ATT GAG GAT GGA ATT TTT GAG GTC AAA TCA ACA      731
Val Ser Ile Leu Thr Ile Glu Asp Gly Ile Phe Glu Val Lys Ser Thr
            210                 215                 220

GCT GGA GAC ACC CAC TTA GGC GGA GAA GAC TTT GAT AAC CGA ATG GTC      779
Ala Gly Asp Thr His Leu Gly Gly Glu Asp Phe Asp Asn Arg Met Val
        225                 230                 235

AAT CAT TTC ATT GCT GAG TTC AAG CGA AAG CAC AAG AAA GAC ATC AGT      827
Asn His Phe Ile Ala Glu Phe Lys Arg Lys His Lys Lys Asp Ile Ser
    240                 245                 250

GAG AAC AAG AGA GCT GTC CGC CGT CTC CGC ACG GCC TGC GAG CGG GCC      875
Glu Asn Lys Arg Ala Val Arg Arg Leu Arg Thr Ala Cys Glu Arg Ala
255                 260                 265                 270

AAG CGC ACC CTC TCC TCC AGC ACC CAG GCC AGT ATT GAG ATT GAT TCT      923
Lys Arg Thr Leu Ser Ser Ser Thr Gln Ala Ser Ile Glu Ile Asp Ser
                275                 280                 285

CTC TAT GAG GGA ATT GAC TTC TAT ACC TCC ATT ACC CGT GCT CGA TTT      971
Leu Tyr Glu Gly Ile Asp Phe Tyr Thr Ser Ile Thr Arg Ala Arg Phe
            290                 295                 300

GAG GAG TTG AAT GCT GAC CTG TTC CGT GGC ACA CTG GAC CCT GTA GAG     1019
Glu Glu Leu Asn Ala Asp Leu Phe Arg Gly Thr Leu Asp Pro Val Glu
        305                 310                 315

AAG GCC CTT CGA GAT GCC AAG CTG GAC AAG TCA CAG ATC CAT GAT ATT     1067
Lys Ala Leu Arg Asp Ala Lys Leu Asp Lys Ser Gln Ile His Asp Ile
    320                 325                 330
```

```
GTC TTG GTG GGT GGT TCT ACC AGA ATC CCC AAG ATC CAG AAA CTT CTG      1115
Val Leu Val Gly Gly Ser Thr Arg Ile Pro Lys Ile Gln Lys Leu Leu
335                 340                 345                 350

CAA GAC TTC TTC AAT GGA AAA GAG CTG AAC AAG AGC ATT AAC CCC GAT      1163
Gln Asp Phe Phe Asn Gly Lys Glu Leu Asn Lys Ser Ile Asn Pro Asp
                355                 360                 365

GAA GCT GTT GCC TAT GGT GCA GCT GTC CAG GCA GCC ATT CTA TCT GGA      1211
Glu Ala Val Ala Tyr Gly Ala Ala Val Gln Ala Ala Ile Leu Ser Gly
            370                 375                 380

GAC AAG TCT GAG AAC GTT CAG GAT TTG CTG CTC TTG GAT GTC ACT CCT      1259
Asp Lys Ser Glu Asn Val Gln Asp Leu Leu Leu Leu Asp Val Thr Pro
        385                 390                 395

CTT TCC CTT GGT ATT GAA ACT GCT GGC GGA GTC ATG ACT GTC CTC ATC      1307
Leu Ser Leu Gly Ile Glu Thr Ala Gly Gly Val Met Thr Val Leu Ile
    400                 405                 410

AAG CGC AAT ACC ACC ATC CCC ACC AAG CAG ACA CAG ACT CTC ACC ACC      1355
Lys Arg Asn Thr Thr Ile Pro Thr Lys Gln Thr Gln Thr Leu Thr Thr
415                 420                 425                 430

TAC TCT GAC AAC CAG CCT GGT GTA CTC ATT CAG GTG TAT GAA GGT GAA      1403
Tyr Ser Asp Asn Gln Pro Gly Val Leu Ile Gln Val Tyr Glu Gly Glu
                435                 440                 445

AGG GCC ATG ACC AAG GAC AAC AAC CTG CTT GGA AAG TTC GAG CTC ACA      1451
Arg Ala Met Thr Lys Asp Asn Asn Leu Leu Gly Lys Phe Glu Leu Thr
            450                 455                 460

GGC ATC CCT CCA GCA CCC CGT GGG GTT CCT CAG ATT GAG GTT ACT TTT      1499
Gly Ile Pro Pro Ala Pro Arg Gly Val Pro Gln Ile Glu Val Thr Phe
        465                 470                 475

GAC ATC GAT GCC AAT GGC ATC CTC AAT GTT TCT GCT GTA GAT AAG AGC      1547
Asp Ile Asp Ala Asn Gly Ile Leu Asn Val Ser Ala Val Asp Lys Ser
    480                 485                 490

ACA GGA AAG GAG AAC AAG ATC ACC ATC ACC AAT GAC AAG GGC CGC TTG      1595
Thr Gly Lys Glu Asn Lys Ile Thr Ile Thr Asn Asp Lys Gly Arg Leu
495                 500                 505                 510

AGT AAG GAA GAT ATT GAG CGC ATG GTC CAA GAA GCT GAG AAG TAC AAG      1643
Ser Lys Glu Asp Ile Glu Arg Met Val Gln Glu Ala Glu Lys Tyr Lys
                515                 520                 525

GCT GAG GAT GAG AAG CAG AGA GAT AAG GTT TCC TCC AAG AAC TCA CTG      1691
Ala Glu Asp Glu Lys Gln Arg Asp Lys Val Ser Ser Lys Asn Ser Leu
            530                 535                 540
```


BNSDOCID <WO 9408012A1 | >

```
GAG TCC TAT GCC TTC AAC ATG AAA GCA ACT GTG GAA GAT GAG AAA CTT      1739
Glu Ser Tyr Ala Phe Asn Met Lys Ala Thr Val Glu Asp Glu Lys Leu
        545                 550                 555

CAA GGC AAG ATC AAT GAT GAG GAC AAA CAG AAG ATT CTT GAC AAG TGC      1787
Gln Gly Lys Ile Asn Asp Glu Asp Lys Gln Lys Ile Leu Asp Lys Cys
    560                 565                 570

AAT GAA ATC ATC AGC TGG CTG GAT AAG AAC CAG ACT GCA GAG AAG GAA      1835
Asn Glu Ile Ile Ser Trp Leu Asp Lys Asn Gln Thr Ala Glu Lys Glu
575                 580                 585                 590

GAA TTT GAG CAT CAG CAG AAA GAA CTG GAG AAA GTC TGC AAC CCT ATT      1883
Glu Phe Glu His Gln Gln Lys Glu Leu Glu Lys Val Cys Asn Pro Ile
                595                 600                 605

ATC ACC AAG CTG TAC CAG AGT GCA GGT GGC ATG CCT GGA GGG ATG CCT      1931
Ile Thr Lys Leu Tyr Gln Ser Ala Gly Gly Met Pro Gly Gly Met Pro
            610                 615                 620

GGT GGC TTC CCA GGT GGA GGA GCT CCC CCA TCT GGT GGT GCT TCT TCA      1979
Gly Gly Phe Pro Gly Gly Gly Ala Pro Pro Ser Gly Gly Ala Ser Ser
        625                 630                 635

GGC CCC ACC ATT GAA GAG GTG GAT TA AGTCAGTCCA AGAAGAAGGT            2025
Gly Pro Thr Ile Glu Glu Val Asp
    640                 645

GTAGCTTTGT TCCACAGGGA CCCAAAAAGT AACATGGAAT AATAAAACTA TTTAAATTGG    2085

CACC                                                                2089
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 646 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Ser Lys Gly Pro Ala Val Gly Ile Asp Leu Gly Thr Thr Tyr Ser
  1               5                  10                  15

Cys Val Gly Val Phe Gln His Gly Lys Val Glu Ile Ile Ala Asn Asp
            20                  25                  30
```


NSDOCID: <WO 9408012A1 | -

Gln Gly Asn Arg Thr Thr Pro Ser Tyr Val Ala Phe Thr Asp Thr Glu
35 40 45

Arg Leu Ile Gly Asp Ala Ala Lys Asn Gln Val Ala Met Asn Pro Thr
50 55 60

Asn Thr Val Phe Asp Ala Lys Arg Leu Ile Gly Arg Arg Phe Asp Asp
65 70 75 80

Ala Val Val Gln Ser Asp Met Lys His Trp Pro Phe Met Val Val Asn
85 90 95

Asp Ala Gly Arg Pro Lys Val Gln Val Glu Tyr Lys Gly Glu Thr Lys
100 105 110

Ser Phe Tyr Pro Glu Glu Val Ser Ser Met Val Leu Thr Lys Met Lys
115 120 125

Glu Ile Ala Glu Ala Tyr Leu Gly Lys Thr Val Thr Asn Ala Val Val
130 135 140

Thr Val Pro Ala Tyr Phe Asn Asp Ser Gln Arg Gln Ala Thr Lys Asp
145 150 155 160

Ala Gly Thr Ile Ala Gly Leu Asn Val Leu Arg Ile Ile Asn Glu Pro
165 170 175

Thr Ala Ala Ala Ile Ala Tyr Gly Leu Asp Lys Lys Val Gly Ala Glu
180 185 190

Arg Asn Val Leu Ile Phe Asp Leu Gly Gly Gly Thr Phe Asp Val Ser
195 200 205

Ile Leu Thr Ile Glu Asp Gly Ile Phe Glu Val Lys Ser Thr Ala Gly
210 215 220

Asp Thr His Leu Gly Gly Glu Asp Phe Asp Asn Arg Met Val Asn His
225 230 235 240

Phe Ile Ala Glu Phe Lys Arg Lys His Lys Lys Asp Ile Ser Glu Asn
245 250 255

Lys Arg Ala Val Arg Arg Leu Arg Thr Ala Cys Glu Arg Ala Lys Arg
260 265 270

Thr Leu Ser Ser Ser Thr Gln Ala Ser Ile Glu Ile Asp Ser Leu Tyr
275 280 285

Glu Gly Ile Asp Phe Tyr Thr Ser Ile Thr Arg Ala Arg Phe Glu Glu
290 295 300

```
Leu Asn Ala Asp Leu Phe Arg Gly Thr Leu Asp Pro Val Glu Lys Ala
305                 310                 315                 320

Leu Arg Asp Ala Lys Leu Asp Lys Ser Gln Ile His Asp Ile Val Leu
                325                 330                 335

Val Gly Gly Ser Thr Arg Ile Pro Lys Ile Gln Lys Leu Leu Gln Asp
            340                 345                 350

Phe Phe Asn Gly Lys Glu Leu Asn Lys Ser Ile Asn Pro Asp Glu Ala
        355                 360                 365

Val Ala Tyr Gly Ala Ala Val Gln Ala Ala Ile Leu Ser Gly Asp Lys
    370                 375                 380

Ser Glu Asn Val Gln Asp Leu Leu Leu Leu Asp Val Thr Pro Leu Ser
385                 390                 395                 400

Leu Gly Ile Glu Thr Ala Gly Gly Val Met Thr Val Leu Ile Lys Arg
                405                 410                 415

Asn Thr Thr Ile Pro Thr Lys Gln Thr Gln Thr Leu Thr Thr Tyr Ser
            420                 425                 430

Asp Asn Gln Pro Gly Val Leu Ile Gln Val Tyr Glu Gly Glu Arg Ala
        435                 440                 445

Met Thr Lys Asp Asn Asn Leu Leu Gly Lys Phe Glu Leu Thr Gly Ile
    450                 455                 460

Pro Pro Ala Pro Arg Gly Val Pro Gln Ile Glu Val Thr Phe Asp Ile
465                 470                 475                 480

Asp Ala Asn Gly Ile Leu Asn Val Ser Ala Val Asp Lys Ser Thr Gly
                485                 490                 495

Lys Glu Asn Lys Ile Thr Ile Thr Asn Asp Lys Gly Arg Leu Ser Lys
            500                 505                 510

Glu Asp Ile Glu Arg Met Val Gln Glu Ala Glu Lys Tyr Lys Ala Glu
        515                 520                 525

Asp Glu Lys Gln Arg Asp Lys Val Ser Ser Lys Asn Ser Leu Glu Ser
    530                 535                 540

Tyr Ala Phe Asn Met Lys Ala Thr Val Glu Asp Glu Lys Leu Gln Gly
545                 550                 555                 560

Lys Ile Asn Asp Glu Asp Lys Gln Lys Ile Leu Asp Lys Cys Asn Glu
                565                 570                 575
```

Ile Ile Ser Trp Leu Asp Lys Asn Gln Thr Ala Glu Lys Glu Glu Phe
580 585 590

Glu His Gln Gln Lys Glu Leu Glu Lys Val Cys Asn Pro Ile Ile Thr
595 600 605

Lys Leu Tyr Gln Ser Ala Gly Gly Met Pro Gly Gly Met Pro Gly Gly
610 615 620

Phe Pro Gly Gly Gly Ala Pro Pro Ser Gly Gly Ala Ser Ser Gly Pro
625 630 635 640

Thr Ile Glu Glu Val Asp
645

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5408 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 1040..1244

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 1569..1772

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 2097..2249

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 2337..2892

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 3104..3306

(ix) FEATURE:

BNSDOCID: <WO 9408012A1 | >

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 3881..4113

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 4445..4629

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGCTTGAAA GTTCCAGAAC GCTGCGGTGA GTGCGTTATC GTGAGGCGGC GCGGTGGGGT 60

GGGTGCGGAA GGGGGCGAGG CGAGGAGTGG AGCCGCGTTG TGATTGTGAT TGGGTCTTGT 120

AAGGGCAGCC GGACTCTATT GGCCGGGAAC CTAATGCAGG AAGCAGGCGG ACCCCTTCTG 180

GAAGGTTCTA AGATAGGGTA TAAGAGGCAG GGTGGCGGGC GGAAACCGGT GCTCAGTTGA 240

ACTGCGCTGC AGCTCTTGGT TTTTTGTGGC TTCCTTCGTT ATTGGAGCCA GGCCTACACC 300

CCAGGTAAAA CCTCTGCTCA AGAGTTGGGT TGTGGGTCTG GGAGCGTGCA GCCTCCACAC 360

AGGCCTGTTG GGCTTGCTGA GGCTTGGGGG TTCTGAGAAT CTCGTCGAGG CGAGTGTGCG 420

GCTCCTTCTA CCGGCTTAAA GGGCCTCAGT TTTCGGTGGG ATGGCAGCGG TATTTGGTTG 480

CAGCCGGCAG ACGGAAATGT AGGGAGTGGG CCGCATGGCC CCAGGGGAGG CTGGGAGACG 540

CCCGGCCGCG TGGCGGGGGA GGGTTGCTGC ATCGGTTTGC CTGGCGCGCG GGGAAGTGGA 600

GCCAGCGTTT TCTTTCACCC AGTTCCCTGC TTAGTCCAGT CCCACCGTGG TTCTTCAGAG 660

CTGTTCTTGG CGTGCTTCCA GTATGGGGGT ACATTCCGGA GTAGTTAAAA GCCCGTTGAC 720

TCCCGGGGGG CACTGGCACC TGGCGAGGGA GGGGAACAGA CAGTGCTCAG TTCGGGGTAA 780

GACCACGTGT TGAGCAACGC CCCACGCCGT CTGGGTCGAT GGGTCCTTCA TCTAGGGCGT 840

GCTGTGCTGC GGTTGGCACG GCAACCTGGA CTGCAGCACT AGTTCTGGAC CTCGCGCGTG 900

CTTAGACAGG AGGTGATGGG CACTATTACC TCTTGGCAGT GGCCATACGT TTTTCCTGGT 960

TAAGTGTTCT GTTAAGGGAT GAGGGAAATA TTTTGATTAA TTGAATTTTT AAACCAGATT 1020

TTTCTTTTTT TCAGCAACCA TGTCCAAGGG ACCTGCAGTT GGTATTGATC TTGGCACCAC 1080

CTACTCTTGT GTGGGTGTTT TCCAGCACGG AAAAGTCGAG ATAATTGCCA ATGATCAGGG 1140

AAACCGAACC ACTCCAAGCT ATGTCGCCTT TACGGACACT GAACGGTTGA TCGGTGATGC 1200

CGCAAAGAAT CAAGTTGCAA TGAACCCCAC CAACACAGTT TTTGGTGAGT TCCTAATTTT 1260

AAATGACAGA ACAAATATAA ACAGGGCTAG CAAGCACAAA AGTTTATGAA ACGTGAGGAG 1320

GGAACTTTTT GATTTTAGAA AAACTGAGCT GAGAGACTTG TTATCAAGTC TGTTATAAAA 1380

CAGGTTGTAG AAACCTTTCA GGCTGAAATC TGGATAACGT AGGAGGTTGA AGTTTGAACC 1440

TTTGCTAGGT ATATGGTAGT TGAATTCACC TACCTATGAA CTGTTAGGTA TTTGAGTAAT 1500

CATGGACTTG AGTTTTATCT GAAGAGCTAT GAAATTGAAA GTGTTTTCAT TTGACACCTT 1560

TTACAGATGC CAAACGTCTG ATTGGACGCA GATTTGATGA TGCTGTTGTC CAGTCTGATA 1620

TGAAACATTG GCCCTTTATG GTGGTGAATG ATGCTGGCAG GCCCAAGGTC CAAGTAGAAT 1680

ACAAGGGAGA GACCAAAAGC TTCTATCCAG AGGAGGTGTC TTCTATGGTT CTGACAAAGA 1740

TGAAGGAAAT TGCAGAAGCC TACCTTGGGA AGGTGAGGTT GGTTTTTCAG TATGGGGTGC 1800

ATTCCGGAGT AGTTAAAAGC CCGATGACTC CCGGGGGCAC TGGCACCTGG CGAGGGAGGG 1860

GAACAGATGG GGCTCAGCTC AGGGTTAAGA CCACGTGCCC AACAGTGCCC TAGGCTCTCT 1920

AGGTAGATGG GTCTGTCAAC ACCAGAAACC AGTGAATCTT GACAATTACA CAGTAATTTA 1980

CATTTTGGTG GGGGGGGTGC TCCAGCTGTT GTTTCACCAG CATTAATCCA TTTGCTGGAG 2040

TTTGCATATA TGTAAGTATA ATAGTTACCA ATCTGTGGTC TTTTCCTTAT TCCTAGACTG 2100

TTACCAATGC TGTGGTCACA GTGCCAGCTT ACTTTAATGA CTCTCAGCGT CAGGCTACCA 2160

AAGATGCTGG AACTATTGCT GGTCTCAATG TACTTAGAAT TATTAATGAG CCAACTGCTG 2220

CTGCTATTGC TTACGGCTTA GACAAAAAGG TATGTACCAT TTGTGATGCA AGTTCGGATT 2280

ATTTTAAGAT TAATTTGATC CATCGTAAAT TTAAATGAGA TTGTTTTTAA CGGCAGGTTG 2340

GAGCAGAAAG AAACGTGCTC ATCTTTGACC TGGGAGGTGG CACTTTTGAT GTGTCAATCC 2400

TCACTATTGA GGATGGAATC TTTGAGGTCA AGTCTACAGC TGGAGACACC CACTTGGGTG 2460

GAGAAGATTT TGACAACCGA ATGGTCAACC ATTTTATTGC TGAGTTTAAG CGCAAGCATA 2520

AGAAGGACAT CAGTGAGAAC AAGAGAGCTG TAAGACGCCT CCGTACTGCT TGTGAACGTG 2580

CTAAGCGTAC CCTCTCTTCC AGCACCCAGG CCAGTATTGA GATCGATTCT CTCTATGAAG 2640

GAATCGACTT CTATACCTCC ATTACCCGTG CCCGATTTGA AGAACTGAAT GCTGACCTGT 2700

TCCGTGGCAC CCTGGACCCA GTAGAGAAAG CCCTTCGAGA TGCCAAACTA GACAAGTCAC 2760

BNSDOCID: <WO 9409012A1 | >

AGATTCATGA TATTGTCCTG GTTGGTGGTT CTACTCGTAT CCCCAAGATT CAGAAGCTTC 2820

TCCAAGACTT CTTCAATGGA AAAGAACTGA ATAAGAGCAT CAACCCTGAT GAAGCTGTTG 2880

CTTATGGTGC AGGTAACAAT GGTATCTCAA TTAACCCTAA AGGCAGGCAG GCCCAAGGTG 2940

ACTCGCTGTG ATGAGTGATT GTTAAACATT CGTAGTTTCC ACCAAAAGCT TGGCTAATGA 3000

TGGCAACACC TTCCTTGGAT GTCTGAGCGA GTGATAGTTA AAACAGGAGC TATGTACTGG 3060

GTTTTCTTTT AACTTCTTTT AACGTTAACT TTTTGTTTGC TAGCTGTCCA GGCAGCCATC 3120

TTGTCTGGAG ACAAGTCTGA GAATGTTCAA GATTTGCTGC TCTTGGATGT CACTCCTCTT 3180

TCCCTTGGTA TTGAAACTGC TGGTGGAGTC ATGACTGTCC TCATCAAGCG TAATACCACC 3240

ATTCCTACCA AGCAGACACA GACCTTCACT ACCTATTCTG ACAACCAGCC TGGTGTGCTT 3300

ATTCAGGTAT GTTTCTGTAC TTCTCTTGTT TGGCTTACTG ATAACAGATA AAGGGAAGTC 3360

TTGACTGACT CGCTATGATG ATGGATTCCA AAACCATTCG TAGTTTCCAC CAGAAAGTCT 3420

TATGTTGGCC AGTTCCTTCC TTGGATGTTT GAGCGACCAT TCTTCCTTAG CAGGACCCTA 3480

GCACTGTCAC AGACCTGGAG TCCATTGTAG TAATTTGTTT TATTTCCTAC CAAGGTTTAT 3540

GAAGGCGAGC GTGCCATGAC AAAGGATAAC AACCTGCTTG GCAAGTTTGA ACTCACAGGC 3600

ATACCTCCTG CACCCCGAGG TGTTCCTCAG ATTGAAGTCA CTTTTGACAT TGATGCCAAT 3660

GGTATACTCA ATGTCTCTGC TGTGGACAAG AGTACGGGAA AAGAGAACAA GATTACTATC 3720

ACTAATGACA AGGGTAAGGA GGCACTGTCA TCTGGTCTTG ACAGGGATAA TGGTATTTCA 3780

ATTGAGTTAC TGGTGAATAA GGGCGTCTAG CTAAGAGAAA CTAGAGTTAC ACATACACAG 3840

GTAATTTAAG GCTTTTACTT AGAGTTAATT TCTTTCCTAG GCCGTTTGAG CAAGGAAGAC 3900

ATTGAACGTA TGGTCCAGGA AGCTGAGAAG TACAAAGCTG AAGATGAGAA GCAGAGGGAC 3960

AAGGTGTCAT CCAAGAATTC ACTTGAGTCC TATGCCTTCA ACATGAAAGC AACTGTTGAA 4020

GATGAGAAAC TTCAAGGCAA GATTAACGAT GAGGACAAAC AGAAGATTCT GGACAAGTGT 4080

AATGAAATTA TCAACTGGCT TGATAAGAAT CAGGTTTGTG TTTTTTTTTT TTTTTTTCCT 4140

CCCCCACGCA ATGGAGGGGA AGGGGATGGT AAACCAAGCT TGAGCTGGAT TTCAGTGTAG 4200

GGTCACAATG ATGAATGGTC CAAAACATTC GCGGTTTCCA CCAGAATTCA AGGTGTTGGC 4260

AACTACCTTC CTTGGATGTC TGAGTGACCC AAGATGTTAA GGAAGAATAA GGCCCTATTT 4320

TAATGTTGGT ATGGGCCCTC TTGTAAGAGT TTGCTCCAGA CTTTTAGTAT CAGATTGCGT 4380

CAGGGAGAAA GAAGGGTTAT TAACATTAAA AGAACTTGCA GTAATTCCTT TTTCTCTTCC 4440

TCAGACTGCT GAGAAGGAAG AATTTGAACA TCAACAGAAA GAGCTGGAGA AAGTTTGCAA 4500

CCCCATCATC ACCAAGCTGT ACCAGAGTGC AGGAGGCATG CCAGGAGGAA TGCCTGGGGG 4560

ATTTCCTGGT GGTGGAGCTC CTCCCTCTGG TGGTGCTTCC TCAGGGCCCA CCATTGAAGA 4620

GGTTGATTAA GCCAACCAAG TGTAGATGTA GCATTGTTCC ACACATTTAA AACATTTGAA 4680

GGACCTAAAT TCGTAGCAAA TTCTGTGGCA GTTTTAAAAA GTTAAGCTGC TATAGTAAGT 4740

TACTGGGCAT TCTCAATACT TGAATATGGA ACATATGCAC AGGGGAAGGA AATAACATTG 4800

CACTTTATAC ACTGTATTGT AAGTGGAAAA TGCAATGTCT TAAATAAAAC TATTTAAAAT 4860

TGGCACCATA CAATTGCTTT GAGTCTTTAA ATAATCTCCC AGGCCAGCGG TGGGAGAAGT 4920

AGGCTTAGGT GATTATGTGA CTCTTACTTT CTCCTTCCTC TTAAGCTTGA GTTAACAAGG 4980

GCTGGGTGGC AAGTTGCCCT TCAGAGCATG TGGATGGTAC ATTTTGGAAT TCAGAGCTTT 5040

GAGAAGGGGA GCATAAGAAA TTGGATCTGG ATCAAACTAA CCTTAGTCCT TAGGCTGGAG 5100

AGGCAGAAGC TGACTTAATG GTGTTTTCTA AACTTATTCT GTGTGTAAGC CTGCCTAGGA 5160

GCAGAGGCTT TCCTGGAGGG TTGTGCTAGA TGAGTAAGAA TTTAGATACA GAATCAAATA 5220

ATGGGCAGTG AATATTAAGC TACATGGCAG AGGTATCTGA ATGTCAATCC CTTATATGAG 5280

CCACTGCCCT GTGGGCTTCC ATTTCTTCTG AGTTAAGATT ATTCAGAAGG TCGGGGATTG 5340

GAGCTAAGCT GCCACCTGGT TAATTAAGGT CCCAACAGTG AGTTGTGATA GCCTAGGGGA 5400

GCAGGCTG 5408

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 666 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

INSDOCID: <AP> 0600012A1 >

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Glu Thr Arg Arg Phe Val Cys Asp Glu Arg Arg Ala Gly Gly Met Arg
1               5                  10                  15

His Leu Leu Leu Ala Leu Leu Leu Leu Gly Gly Ala Arg Ala Asp Asp
            20                  25                  30

Glu Glu Lys Lys Glu Asp Val Gly Thr Val Val Gly Ile Asp Leu Gly
        35                  40                  45

Thr Thr Tyr Ser Cys Val Gly Val Phe Lys Asn Gly Arg Val Glu Ile
    50                  55                  60

Ile Ala Asn Asp Gln Gly Asn Arg Ile Thr Pro Ser Tyr Val Ala Phe
65                  70                  75                  80

Thr Pro Glu Gly Glu Arg Leu Ile Gly Asp Ala Ala Lys Asn Gln Leu
                85                  90                  95

Thr Ser Asn Pro Glu Asn Thr Val Phe Asp Ala Lys Arg Leu Ile Gly
            100                 105                 110

Arg Thr Trp Asn Asp Pro Ser Val Gln Gln Asp Ile Lys Tyr Leu Pro
        115                 120                 125

Phe Lys Val Val Glu Lys Lys Ala Lys Pro His Ile Gln Val Asp Val
    130                 135                 140

Gly Gly Gly Gln Thr Lys Thr Phe Ala Pro Glu Glu Ile Ser Ala Met
145                 150                 155                 160

Val Leu Thr Lys Met Lys Glu Thr Ala Glu Ala Tyr Leu Gly Lys Lys
                165                 170                 175

Val Thr His Ala Val Val Thr Val Pro Ala Tyr Phe Asn Asp Ala Gln
            180                 185                 190

Arg Gln Ala Thr Lys Asp Ala Gly Thr Ile Ala Gly Leu Asn Val Met
        195                 200                 205

Arg Ile Ile Asn Glu Pro Thr Ala Ala Ala Ile Ala Tyr Gly Leu Asp
    210                 215                 220

Lys Arg Glu Gly Glu Lys Asn Ile Leu Val Phe Asp Leu Gly Gly Gly
225                 230                 235                 240

Thr Phe Asp Val Ser Leu Leu Thr Ile Asp Asn Gly Val Phe Glu Val
                245                 250                 255
```


BNSDOCID <WO 9408012A1 | >

Val Ala Thr Asn Gly Asp Thr His Leu Gly Gly Glu Asp Phe Asp Gln
260 265 270

Arg Val Met Glu His Phe Ile Lys Leu Tyr Lys Lys Lys Thr Gly Lys
275 280 285

Asp Val Arg Lys Asp Asn Arg Ala Val Gln Lys Leu Arg Arg Glu Val
290 295 300

Glu Lys Ala Lys Arg Ala Leu Ser Ser Gln His Gln Ala Arg Ile Glu
305 310 315 320

Ile Glu Ser Phe Phe Glu Gly Glu Asp Phe Ser Glu Thr Leu Thr Arg
325 330 335

Ala Lys Phe Glu Glu Leu Asn Met Asp Leu Phe Arg Ser Thr Met Lys
340 345 350

Pro Val Gln Lys Val Leu Glu Asp Ser Asp Leu Lys Lys Ser Asp Ile
355 360 365

Asp Glu Ile Val Leu Val Gly Gly Ser Thr Arg Ile Pro Lys Ile Gln
370 375 380

Gln Leu Val Lys Glu Phe Phe Asn Gly Lys Glu Pro Ser Arg Gly Ile
385 390 395 400

Asn Pro Asp Glu Ala Val Ala Tyr Gly Ala Ala Val Gln Ala Gly Val
405 410 415

Leu Ser Gly Asp Gln Asp Thr Gly Asp Leu Val Leu Leu Asp Val Cys
420 425 430

Pro Leu Thr Leu Gly Ile Glu Thr Val Gly Gly Val Met Thr Lys Leu
435 440 445

Ile Pro Arg Asn Thr Val Val Pro Thr Lys Lys Ser Gln Ile Phe Ser
450 455 460

Thr Ala Ser Asp Asn Gln Pro Thr Val Thr Ile Lys Val Tyr Glu Gly
465 470 475 480

Glu Arg Pro Leu Thr Lys Asp Asn His Leu Leu Gly Thr Phe Asp Leu
485 490 495

Thr Gly Ile Pro Pro Ala Pro Arg Gly Val Pro Gln Ile Glu Val Thr
500 505 510

Phe Glu Ile Asp Val Asn Gly Ile Leu Arg Val Thr Ala Glu Asp Lys
515 520 525

Gly Thr Gly Asn Lys Asn Lys Ile Thr Ile Thr Asn Asp Gln Asn Arg
530 535 540

Leu Thr Pro Glu Glu Ile Glu Arg Met Val Asn Asp Ala Glu Lys Phe
545 550 555 560

Ala Glu Glu Asp Lys Lys Leu Lys Glu Arg Ile Asp Ala Arg Asn Glu
565 570 575

Leu Glu Ser Tyr Ala Tyr Ser Leu Lys Asn Gln Ile Gly Asp Lys Glu
580 585 590

Lys Leu Gly Gly Lys Leu Ser Ser Glu Asp Lys Glu Thr Ile Glu Lys
595 600 605

Ala Val Glu Glu Lys Ile Glu Trp Leu Glu Ser His Gln Asp Ala Asp
610 615 620

Ile Glu Asp Phe Lys Ser Lys Lys Lys Glu Leu Glu Glu Val Val Gln
625 630 635 640

Pro Ile Val Ser Lys Leu Tyr Gly Ser Ala Gly Pro Pro Pro Thr Gly
645 650 655

Glu Glu Glu Ala Ala Glu Lys Asp Glu Leu
660 665

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AAGGGGTTGA CCGTCCGTCG GCACACCACT TATAATGCGG GGTGCAAGCC CCCCGTCTAA 60

AATTTTTTTT TTTTCCATTT TTGTCGTTAT TGTTATTTCC CGTTTTTTGT TTTTTTTGAT 120

TTTTTCGGAG CGACAAACCT TTCGAAACAC GTGTCCTGAA AATTATCCTG GGCTGCACGT 180

GATAATATGT TACCCTGTCG GGCGGCGCCT CTTTTTCCCT TTTCTCTCAC TAGTCTCTTT 240

TTCCAATTTG CCACCGTGTA GCATTTTGTT GTGCTGTTAC AACCACAACA AAACGAAAAA 300

CCCGTATGGA CATACATATA TATATATATA TATATATATA TATATTTTGT TACGCGTGCA 360

TTTTCTTGTT GCAAGCAGCA TGTCTAATTG GTAATTTTAA AGCTGCCAAG CTCTACATAA 420

AGAAAAACAT ACATCTATCC CGTTATGAAG TTTTCTGCTG GTGCCGTCCT GTCATGGTCC 480

TCCCTGCTGC TCGCCTCCTC TGTTTTCGCC CAACAAGAGG CTGTGGCCCC TGAAGACTCC 540

GCTGTCGTTA AGTTGGCCAC CGACTCTTTC AATGAATACA TTCAGTCGCA CGACTTGGTG 600

CTTGCGGAGT TTTTTGCTCC ATGGTGTGGC CACTGTAAGA ACATGGCTCC TGAATACGTT 660

AAAGCCGCCG AGACTTTAGT TGAGAAAAAC ATTACCTTGG CCCAGATCGA CTGTACTGAA 720

AACCAGGATC TGTGTATGGA ACACAACATT CCAGGGTTCC CAAGCTTGAA GATTTTCAAA 780

AACAGCGATG TTAACAACTC GATCGATTAC GAGGGACCTA GAACTGCCGA GGCCATTGTC 840

CAATTCATGA TCAAGCAAAG CCAACCGGCT GTCGCCGTTG TTGCTGATCT ACCAGCTTAC 900

CTTGCTAACG AGACTTTTGT CACTCCAGTT ATCGTCCAAT CCGGTAAGAT TGACGCCGAC 960

TTCAACGCCA CCTTTTACTC CATGGCCAAC AAACACTTCA ACGACTACGA CTTTGTCTCC 1020

GCTGAAAACG CAGACGATGA TTTCAAGCTT TCTATTTACT TGCCCTCCGC CATGGACGAG 1080

CCTGTAGTAT ACAACGGTAA GAAAGCCGAT ATCGCTGACG CTGATGTTTT TGAAAAATGG 1140

TTGCAAGTGG AAGCCTTGCC CTACTTTGGT GAAATCGACG GTTCCGTTTT CGCCCAATAC 1200

GTCGAAAGCG GTTTGCCTTT GGGTTACTTG TTCTACAATG ACGAGGAAGA ATTGGAAGAT 1260

TACAAGCCTC TCTTTACCGA GTTGGCCAAA AAGAACAGAG GTCTAATGAA CTTTGTTAGC 1320

ATCGATGCCA GAAAATTCGG CAGACACGCC GGCAACTTGA ACATGAAGGA ACAATTCCCT 1380

CTATTTGCCA TCCACGACAT GACTGAAGAC TTGAAGTACG GTTTGCCTCA ACTCTCTGAA 1440

GAGGCGTTTG ACGAATTGAG CGACAAGATC GTGTTGGAGT CCAAGGCTAT TGAATCTTTG 1500

GTTAAGGACT TCTTGAAAGG TGATGCCTCC CCAATCGTGA AGTCCCAAGA GATCTTCGAG 1560

AACCAAGATT CCTCTGTCTT CCAATTGGTC GGTAAGAACC ATGACGAAAT CGTCAACGAC 1620

CCAAAGAAGG ACGTTCTTGT TTTGTACTAT GCCCCATGGT GTGGTCACTG TAAGAGATTG 1680

GCCCCAACTT ACCAAGAACT AGCTGATACC TACGCCAACG CCACAACCGA CGTTTTGATT 1740

GCTAAACTAG ACCACACTGA AAACGATGTC AGAGGCGTCG TAATTGAAGG TTACCCAACA 1800

ATCGTCTTAT ACCCAGGTGG TAAGAAGTCC GAATCTGTTG TGTACCAAGG TTCAAGATCC 1860

TTGGACTCTT TATTCGACTT CATCAAGGAA AACGGTCACT TCGACGTCGA CGGTAAGGCC 1920

TTGTACGAAG AAGCCCAGGA AAAAGCTGCT GAGGAAGCCG ATGCTGACGC TGAATTGGCT 1980

GACGAAGAAG ATGCCATTCA CGATGAATTG TAATTCTGAT CACTTTGGTT TTTCATTAAA 2040

TAGAGATATA TAAGAAATTT TCTAGGAAGT TTTTTTAAAA AAAATCATAA AAAGATAAAC 2100

GTTAAAATTC AAACACAATA GTCGTTCGCT ATATTCGTCA CACTGCACGA ACGCCTTAGG 2160

GAAAGAGAAA ATTGACCACG TAGTAATAAT AAGTGCATGG CATCGTCTTT TACTTAAATG 2220

TGGACACTTG CTTTACTGCT TAGGAAACTA CTTATCTCAT CCTCCTCCAT TCCCCTCCCT 2280

TTTCCAATTA CCGTAATAAA AGATGGCTGT ATTTACTCCT CCATCAGGTA ATAGCAATTC 2340

CGACCATACT CACACACAAG ATGACCACGA CAAAGATGAT ATGATATCAA GAAATTCTAT 2400

ACA 2403

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 504 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Lys Phe Ser Ala Gly Ala Val Leu Ser Trp Ser Ser Leu Leu Leu
1 5 10 15

Ala Ser Ser Val Phe Ala Gln Gln Glu Ala Val Ala Pro Glu Asp Ser
20 25 30

Ala Val Val Lys Leu Ala Thr Asp Ser Phe Asn Glu Tyr Ile Gln Ser
35 40 45

His Asp Leu Val Lys Ala Ala Glu Thr Leu Val Glu Lys Asn Ile Thr
50 55 60

Leu Ala Gln Ile Asp Cys Thr Glu Asn Gln Asp Leu Cys Met Glu His
65 70 75 80

Asn Ile Pro Gly Phe Pro Ser Leu Lys Ile Phe Lys Asn Ser Asp Val
85 90 95

```
Asn Asn Ser Ile Asp Tyr Glu Gly Pro Arg Thr Ala Glu Ala Ile Val
            100                 105                 110

Gln Pro Met Ile Lys Gln Ser Gln Pro Ala Val Ala Val Val Ala Val
        115                 120                 125

Val Ala Asp Leu Pro Ala Tyr Leu Ala Asn Glu Thr Phe Val Thr Pro
    130                 135                 140

Val Ile Val Gln Ser Gly Lys Ile Asp Ala Asp Phe Asn Ala Thr Phe
145                 150                 155                 160

Tyr Ser Met Ala Asn Lys His Phe Asn Asp Tyr Asp Phe Val Ser Ala
                165                 170                 175

Glu Asn Ala Asp Asp Asp Phe Lys Leu Ser Ile Tyr Leu Pro Ser Ala
            180                 185                 190

Met Asp Glu Pro Val Val Tyr Asn Gly Lys Lys Ala Asp Ile Ala Asp
        195                 200                 205

Ala Asp Val Phe Glu Lys Trp Leu Gln Val Glu Ala Leu Pro Tyr Phe
    210                 215                 220

Gly Glu Ile Asp Gly Ser Val Phe Ala Gln Tyr Val Glu Ser Gly Leu
225                 230                 235                 240

Pro Leu Gly Tyr Leu Phe Tyr Asn Asp Glu Glu Glu Leu Glu Glu Tyr
                245                 250                 255

Lys Pro Leu Phe Thr Glu Leu Ala Lys Lys Asn Arg Gly Leu Met Asn
            260                 265                 270

Phe Val Ser Ile Asp Ala Arg Lys Phe Gly Arg His Ala Gly Asn Leu
        275                 280                 285

Asn Met Lys Glu Gln Phe Pro Leu Phe Ala Ile His Asp Met Thr Glu
    290                 295                 300

Asp Leu Lys Tyr Gly Leu Pro Gln Leu Ser Glu Glu Ala Phe Asp Glu
305                 310                 315                 320

Leu Ser Asp Lys Ile Val Leu Glu Ser Lys Ala Ile Glu Ser Leu Val
                325                 330                 335

Lys Asp Phe Leu Lys Gly Asp Ala Ser Pro Ile Val Lys Ser Gln Glu
            340                 345                 350

Ile Phe Glu Asn Gln Asp Ser Ser Val Phe Gln Leu Val Gly Lys Asn
        355                 360                 365
```


BNSDOCID: <WO 9408012A1 >

His Asp Glu Ile Val Asn Asp Pro Lys Lys Asp Val Leu Val Leu Tyr
370 375 380

Ala Pro Trp Cys Gly His Cys Lys Arg Leu Ala Pro Thr Tyr Gln Glu
385 390 395 400

Leu Ala Asp Thr Tyr Ala Asn Ala Thr Ser Asp Val Leu Ile Ala Lys
405 410 415

Leu Asp His Thr Glu Asn Asp Val Arg Gly Val Val Ile Glu Gly Tyr
420 425 430

Pro Thr Ile Val Leu Tyr Pro Gly Gly Lys Lys Ser Glu Ser Val Val
435 440 445

Tyr Gln Gly Ser Arg Ser Leu Asp Ser Leu Phe Asp Pro Ile Lys Glu
450 455 460

Asn Gly His Phe Asp Val Asp Gly Lys Ala Leu Tyr Glu Glu Ala Gln
465 470 475 480

Glu Lys Ala Ala Glu Glu Ala Asp Ala Asp Ala Glu Leu Ala Asp Glu
485 490 495

Glu Asp Ala Ile His Asp Glu Leu
500

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCCCGGCGCC AACCTAGCTG CCCCGCCCGC TGCCGACGTC CGACATGCTG AGCCGTGCTT 60

TGCTGTGCCT GGCCCTGGCC TGGGCGGCTA GGGTGGGCGC CGACGCTCTG GAGGAGGAGG 120

ACAACGTCTC GGTGCTGAAG AAGAGCAACT TCGCAGAGCC GGCGGCGCAC AACTACCTGC 180

TGGTGGAGTT CTATGCCCCA TGGTGTGGCC ACTGCAAAGC ATCGGCCCCA GAGTATGCCA 240

AAGCTGCTGC AAAACTGAAG GCAGAAGGAC TCGAGATCCG ACTAGCAAAG GTGGACGCCA 300

CAGAAGAGTC TGACCTGGCC CAGCAGTATG GTGTCCGTGG CTACCCCACA ATCAAGTTCT 360

TCAAGAATGG AGACACAGCC TCCCCAAAGG AATATACAGC TGCCACGGAA GCTGACGACA 420

TTGTGAACTG GCTGAAGAAA CGCACAGGCC CAGCAGCCAC AACCCTGTCT GACACTGCAG 480

CTGCAGAGTC CTTGCTGGAC TCAAGCGAAG TGACGGCTAT CGGCTTCTTC AAGGACGCAG 540

GGTCAGACTC CGCCAAGCAG TTCTTGCTGG CAGCAGAGGC TGCTGATGAC ATACCTTTTG 600

GAATCACTTC CAATTGCGTG TTTTCCAAGT ACCAGCTGGA CAACGATGGG GTGGTCCTCT 660

TTAAGAAGTT TGATGAAGGC CGCAACAATT TTGAATGGTG AGATCACCAA GGAGAAGCTA 720

TTAGACTTCA TCAAGCACAA CCAGCTGCCT TTGGTCATCG AGTTCACTGA ACAGACAGCT 780

CCAAAGATTT TCGGAGGTGA AATCAAGACA CATATTCTGC TGTTCCTGCC CAAGAGTGTG 840

TCTGACTACG ATGGCAAATT GAGCAACTTT AAGAAAGCGG CCGAGGGCTT TAAGGGCAAG 900

ATCCTGTTCA TCTTCATCGA TAGTGACCAC ACTGACAACC AGCGCATACT TGAGTTCTTT 960

GGCCTGAAGA AGGAGGAATG TCCAGCTGTG CGGCTTATTA CCCTGGAGGA AGAGATGACC 1020

AAGTACAAAC CGGAGTCAGA CGAGCTGACA GCTGAGAAGA TCACACAATT TTGCCACCAC 1080

TTCCTGGAGG GCAAGATCAA GCCCCACCTG ATGAGCCAGG AACTGCCTGA AGACTGGGAC 1140

AAGCAGCCAG TGAAAGTGCT AGTTGGGAAA AACTTTGAGG AGGTTGCTTT TGATGAGAAA 1200

AAGAACGTGT TTGTTGAATT CTATGCTCCC TGGTGTGGTC ACTGCAAGCA GCTAGCCCCG 1260

ATTTGGGATA AACTGGGAGA GACATACAAA GACCATGAGA ATATCGTCAT CGCTAAGATG 1320

GACTCAACAG CCAATGAGGT GGAAGCTGTG AAGCTGCACA CCTTTCCCAC ACTCAAGTTC 1380

TTCCCAGCAA GTGCAGACAG AACGGTCATT GATTACAACG GTCAGCGGAC ACTAGATGGT 1440

TTTAAGAAAT TCTTGGAGAG CGGTGGCCAG GATGGAGCGG GGGACAATGA CGACCTCGAC 1500

CTAGAAGAAG CTTTAGAGCC AGATATGGAA GAAGACGACG ATCAGAAAGC CGTGAAGGAT 1560

GAACTGTAGT CGAGAAGCCA GATCTGGCGC CCTGAACCCA AAACCTCGGT GGGCCATGTC 1620

CCAGCAGCCC ACATCTCCGG AGCCTGAGCC TCACCCCAGG AGGGAGCGCC ATCAGAACCC 1680

AGGGAATCTT TCTGAAGCCA CACTCATCTG ACACACGTAC ACTTAAACCT GTCTCTTCTT 1740

TTTTTGCTTT TCAATTTTGG AAAGGGATCT CTGTCCAGGC CAGCCCATCT TGAAGGGCTA 1800

CGTTTTGTTT TAATTGGTGG TGTACTTTTT TGTACGTGGA TTTTGTCCCA AGTGCTTGCT 1860

BNSDOCID: <WO 9408012A1 >

ACCATATTTG GGGATTTCAC ACTGGTAATG TCTTTCCTGT TAGAGAGGTT TATGCTATCA 1920

CTTCAGATTT CGTCTGTGAG ATCTTTCATC TTCCTGACAT GTTCTCATGT CGAGGTACTT 1980

GTTCCACCAC GCAGATTCCC CTGAGACCCC TTCCTGCCCT GCGCAGGAGG CGATCGTTCT 2040

GGGTCGTATG CTCTCTCTCT CTCCACCTTG TACTAGTGTT GCCATGACAG CTAGGCTTTT 2100

GTAGTTTGCA TTTAACCTGG GGATTTCTGC ATCCTGTCAG AGGCTGGGTC CCCACGTGTG 2160

GAAAAGAGAC AGTGGTGGCT TGCTGCCAGG CACAGGCCAG GCCTGGACAG CTCTCACTCT 2220

TCTTAAGCCA GAACTACCGA CCAGCCGGCC GGCTGTCCGC ACATTACTCT GGCTCCTGGA 2280

TCCTCTTCCA GCATGGCATG TGGCCTGTGT GAGGCAGAAC CGGGACCCTT GATTCCCAGA 2340

CTGGGAGTCA GCTAAGGACA CTGGCGCTGA ATGAAATGCC CATTCTCAAG GTCTATTTCT 2400

AAACCATAAT GTTGGAATTG AACACATTGG CTAAATAAAG TTGAAATTTT ACTACCATAA 2460

AAAAAAAAAA AAA 2473

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 510 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Leu Ser Arg Ala Leu Leu Cys Leu Ala Leu Ala Trp Ala Ala Arg
1 5 10 15

Val Gly Ala Asp Ala Leu Glu Glu Glu Asp Asn Val Leu Val Leu Lys
20 25 30

Lys Ser Asn Phe Ala Glu Pro Ala Ala His Asn Tyr Leu Leu Val Glu
35 40 45

Phe Tyr Ala Pro Trp Cys Gly His Cys Lys Ala Leu Ala Pro Glu Tyr
50 55 60

Ala Lys Ala Ala Ala Lys Leu Lys Ala Glu Gly Ser Glu Ile Arg Leu
65 70 75 80

Ala Lys Val Asp Ala Thr Glu Glu Ser Asp Leu Ala Gln Gln Tyr Gly
85 90 95

Val Arg Gly Tyr Pro Thr Ile Lys Phe Phe Lys Asn Gly Asp Thr Ala
100 105 110

Ser Pro Lys Glu Tyr Thr Ala Gly Arg Glu Ala Asp Asp Ile Val Asn
115 120 125

Trp Leu Lys Lys Arg Thr Gly Pro Ala Ala Thr Thr Leu Ser Asp Thr
130 135 140

Ala Ala Ala Glu Ser Leu Val Asp Ser Ser Glu Val Thr Val Ile Gly
145 150 155 160

Phe Phe Lys Asp Ala Gly Ser Asp Ser Ala Lys Gln Phe Leu Leu Ala
165 170 175

Ala Glu Ala Val Asp Asp Ile Pro Phe Gly Ile Thr Ser Asn Ser Asp
180 185 190

Val Phe Ser Lys Tyr Gln Leu Asp Lys Asp Gly Val Val Leu Phe Lys
195 200 205

Lys Phe Asp Glu Gly Arg Asn Asn Phe Glu Gly Glu Ile Thr Lys Glu
210 215 220

Lys Leu Leu Asp Phe Ile Lys His Asn Gln Leu Pro Leu Val Ile Glu
225 230 235 240

Phe Thr Glu Gln Thr Ala Pro Lys Ile Phe Gly Gly Glu Ile Lys Thr
245 250 255

His Ile Leu Leu Phe Leu Pro Lys Ser Val Ser Asp Tyr Asp Gly Lys
260 265 270

Leu Ser Asn Phe Lys Lys Ala Ala Glu Gly Phe Lys Gly Lys Ile Leu
275 280 285

Phe Ile Phe Ile Asp Ser Asp His Thr Asp Asn Gln Arg Ile Leu Glu
290 295 300

Phe Phe Gly Leu Lys Lys Glu Glu Cys Pro Ala Val Arg Leu Ile Thr
305 310 315 320

Leu Glu Glu Glu Met Thr Lys Tyr Lys Pro Glu Ser Asp Glu Leu Thr
325 330 335

Ala Glu Lys Ile Thr Gln Phe Cys His His Phe Leu Glu Gly Lys Ile
340 345 350

BNSDOCID: <WO 9408012A1 | >

```
Lys Pro His Leu Met Ser Gln Ile Glu Leu Pro Glu Asp Trp Asp Lys
        355                 360                 365

Gln Pro Val Lys Val Leu Val Gly Lys Asn Phe Glu Glu Val Ala Pro
    370                 375                 380

Asp Glu Lys Lys Asn Val Phe Val Glu Phe Tyr Ala Pro Trp Cys Gly
385                 390                 395                 400

His Cys Lys Gln Leu Ala Pro Ile Trp Asp Lys Leu Gly Glu Thr Tyr
                405                 410                 415

Lys Asp His Asp Glu Asn Ile Val Ile Ala Lys Met Asp Ser Thr Ala
            420                 425                 430

Asn Glu Val Glu Ala Val Lys Val His Ser Phe Pro Thr Leu Lys Phe
        435                 440                 445

Phe Pro Ala Ser Ala Asp Arg Thr Val Ile Asp Tyr Asn Gly Glu Arg
    450                 455                 460

Thr Leu Asp Gly Phe Lys Lys Phe Leu Glu Ser Gly Gly Gln Asp Gly
465                 470                 475                 480

Ala Gly Asp Asn Asp Asp Leu Asp Leu Glu Glu Ala Leu Glu Pro Asp
                485                 490                 495

Met Glu Glu Asp Asp Asp Gln Lys Ala Val Lys Asp Glu Leu
            500                 505                 510
```

WHAT IS CLAIMED:

1. A method for increasing secretion of an overexpressed gene product from a host cell which comprises effecting the expression of at least one chaperone protein capable of increasing secretion of said overexpressed gene product in said host cell.

2. The method of Claim 1 wherein said expression of said chaperone protein is effected by inducing expression of a nucleic acid encoding said chaperone protein.

3. The method of Claim 2 wherein said nucleic acid is present in an expression vector.

4. A method for increasing secretion of an overexpressed gene product from a host cell which comprises a) effecting the expression of at least one chaperone protein and the overexpression of a gene product in a host cell; and

b) cultivating said host cell under conditions suitable for secretion of said overexpressed gene product.

5. The method of Claim 4 wherein said expression of said chaperone protein is effected by transforming said host cell with an expression vector comprising a nucleic acid encoding said chaperone protein.

6. The method of Claim 5 wherein said overexpression of said gene product is effected by transforming said host cell with an expression vector comprising a nucleic acid encoding said gene product.

7. The method of any one of Claims 1-6 wherein said chaperone protein is an hsp70 chaperone protein or a protein disulfide isomerase.

8. The method of Claim 7 wherein said hsp70 chaperone protein is a KAR2 or a BiP chaperone protein.

9. The method of Claim 7 wherein said protein disulfide isomerase is a mammalian protein disulfide isomerase or a yeast protein disulfide isomerase.

10. A method for increasing secretion of an overexpressed gene product from a host cell which comprises effecting the expression of an hsp70 chaperone protein and a protein disulfide isomerase protein in said host cell.

11. The method of Claim 10 wherein said host cell is a yeast cell.

12. The method of Claim 11 wherein said hsp70 chaperone protein is KAR2 and said protein disulfide isomerase is yeast protein disulfide isomerase.

13. A method for increasing secretion of an overexpressed gene product which comprises transforming a host cell with an expression vector comprising a nucleic acid encoding said gene product under conditions suitable for expression of said gene product, wherein said host cell is overexpressing at least one chaperone protein.

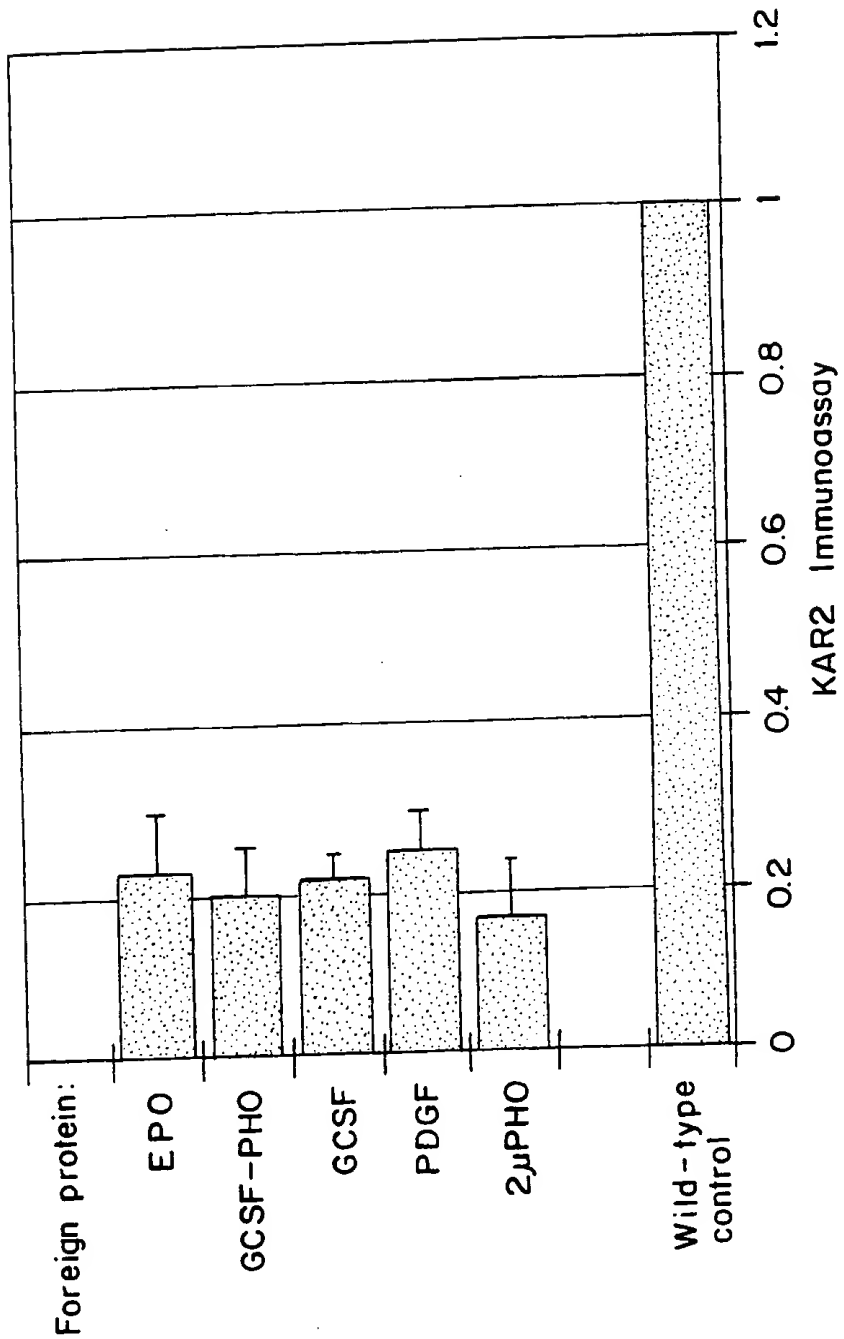
14. The method of Claim 13 wherein said host cell is overexpressing an hsp70 chaperone protein and a protein disulfide isomerase.

15. The method of Claim 13 wherein said chaperone protein is an hsp70 chaperone protein or a protein disulfide isomerase.

16. The method of Claims 14 or 15 wherein said hsp chaperone protein is KAR2 and said protein disulfide isomerase is yeast protein disulfide isomerase.

17. The method of Claim 16 wherein said host cell is a yeast cell.

FIG.1



FINSDOCID: <WO 9408012A1 >

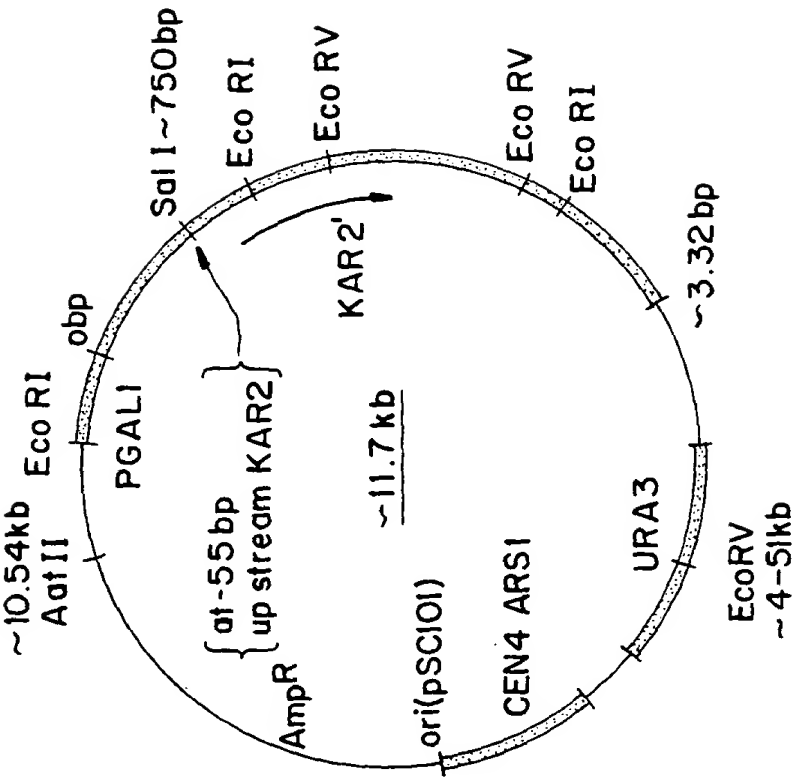


FIG.2

Plasmid Name: pMR1341

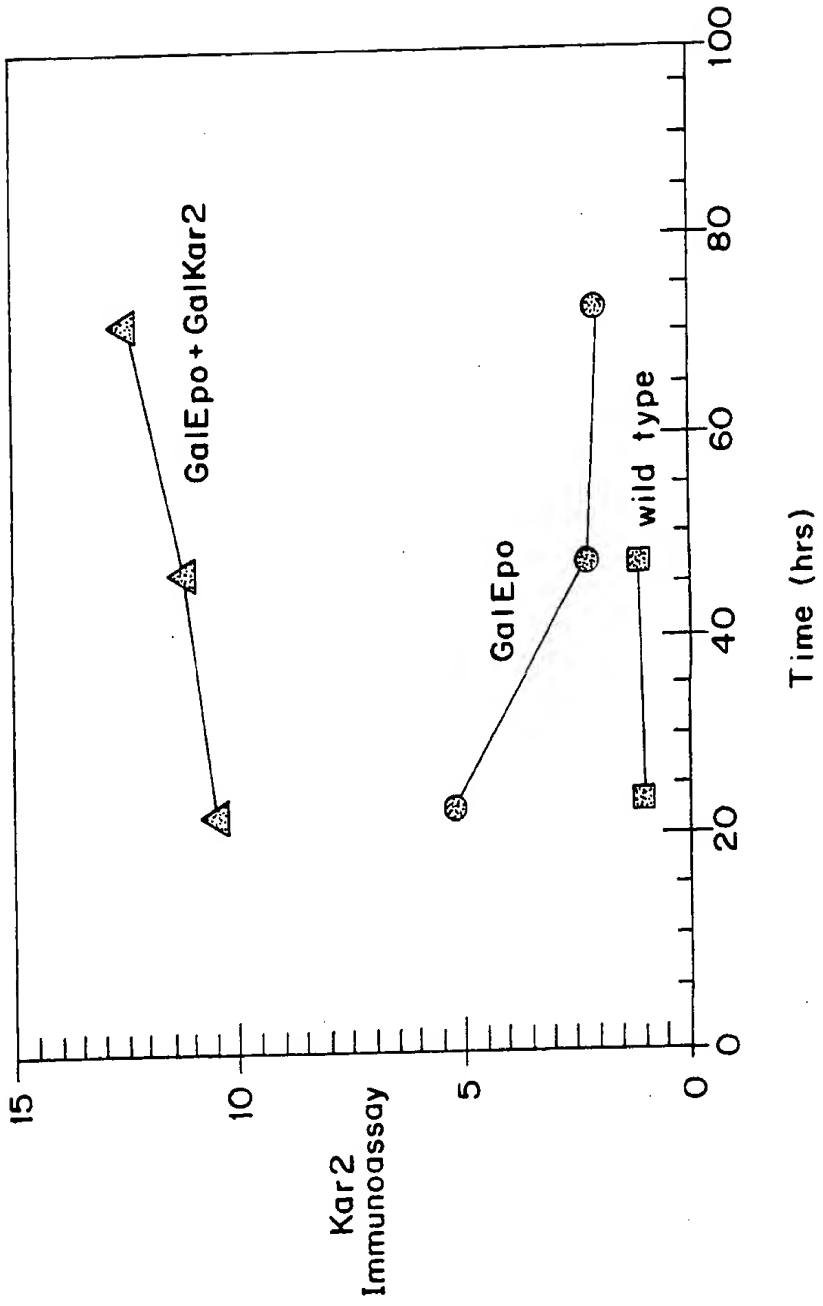
Strain: MR1341

Host and Markers:
HB101

Plasmid Markers:
Amp^R low copy bac. replicon
ARSICEN4
URA3
PGAL1'-KAR2'
Origin:
Sal I/Aat II PGAL1 fragment from pB622
put into Sal I/Aat II sites of pMR568

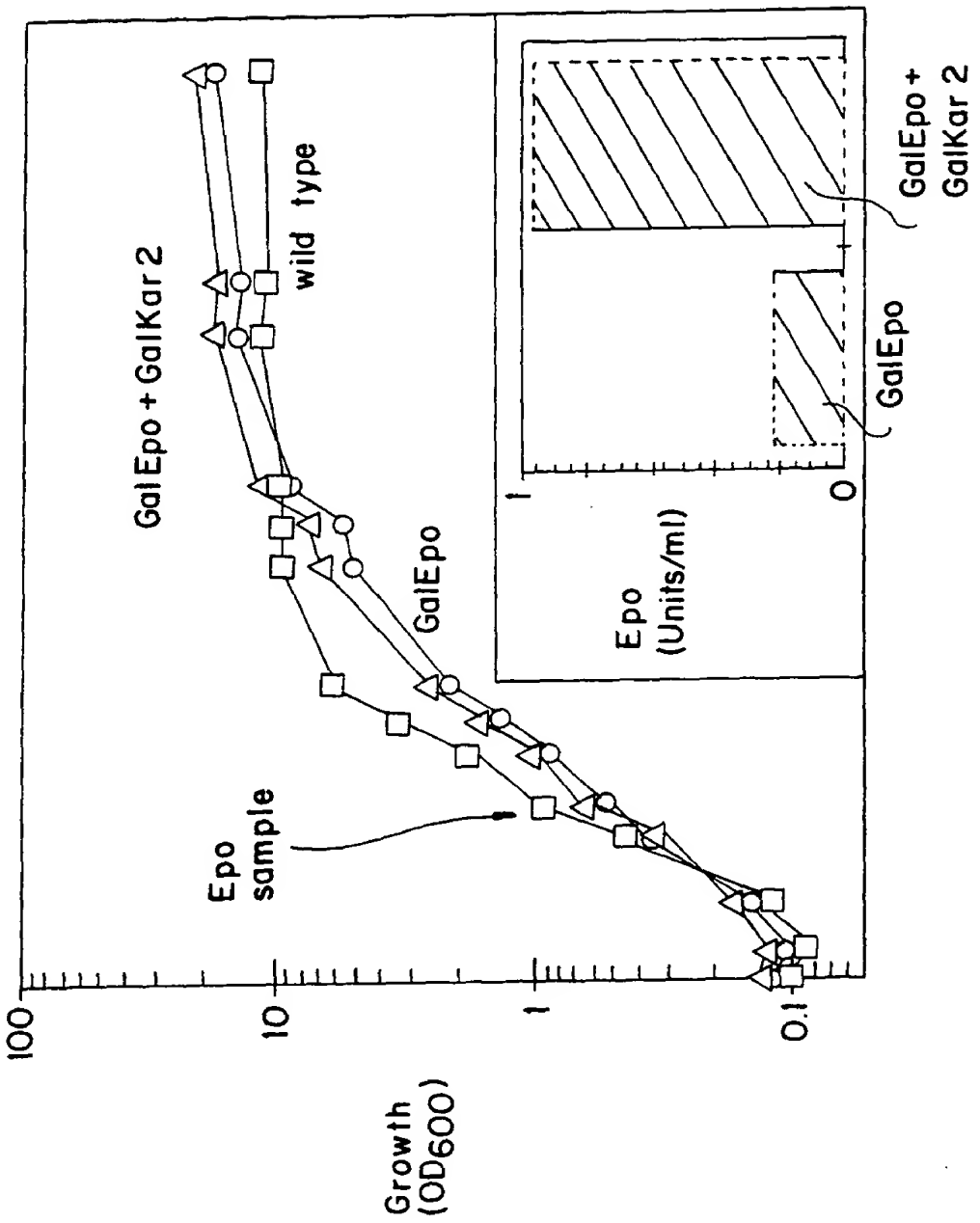
RNSDOCID: WO 9408012A1 I >

FIG.3



BNSDOCID: <WO 9408012A1 | >

FIG.4


GalEpo + GalKar 2
wild type
GalEpo
Epo sample
100
10
1
0.1
Growth (OD600)
GalEpo + GalKar 2
GalEpo
1
0
Epo (Units/ml)

INTERNATIONAL SEARCH REPORT

Inter[illegible]onal Application No
P[illegible]US 93/09426

A. CLASSIFICATION OF SUBJECT MA[illegible]R
IPC 5 C12N15/31 C12N15/12 C12N15/61 C12N15/62 C12N15/81
C12N9/90

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ABSTR. PAP. AM. CHEM. SOC. vol. 203, no. 1-3 , 1992 , ACS,WASHINGTON,DC,US; page BTECH45 A.S. ROBINSON AND K.D. WITTRUP 'Interaction of KAR/BIP with foreign proteins secreted in yeast' 203rd ACS National Meeting, San Francisco, California, April 5-10, 1992; abstract no. 45 ---	1-17
Y	BIOCHEMISTRY; BY D. VOET/ J.G. VOET 1990 , J. WILEY & SONS, INC., US; see page 49, right column, line 34 - page 50, left column, line 9 see page 419, left column, line 40 - line 47 --- -/--	1-17

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 January 1994	08.02.94
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer Hornig, H

Form PCT/ISA/210 (second sheet) (July 1992)

BNSDOCID: <WO 9408012A1_1 >

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EMBO JOURNAL vol. 11, no. 4 , April 1992 , IRL PRESS LIM., OXFORD, ENGL.; pages 1573 - 1581 M.R. KNITTLER AND I.G. HAAS 'Interaction of BiP with newly synthesized immunglobulin light chain molecules: cycles of sequential binding and release' see page 1573, left column, line 1 - line 17 ---	1-17
Y	BIO/TECHNOLOGY vol. 10, no. 6 , June 1992 , NATURE AMERICA, INC., NEW YORK, US; pages 682 - 685 J. BUCHNER ET AL. 'Renaturation of a single-chain immunotoxin facilitated by chaperones and protein disulfide isomerase' see page 682, left column, line 1 - page 683, right column, line 7 see page 684, left column, paragraph 3 see page 685, left column, line 4 - line 46 ---	1-17
Y	J. BIOL. CHEM. vol. 264, no. 34 , 5 December 1989 , AM. SOC. MOL. BIOL., INC.,BALTIMORE, US; pages 20602 - 20607 A.J. DORNER ET AL. 'Increased synthesis of secreted proteins induces expression of glucose-regulated proteins in butyrate-treated chinese hamster ovary cells' see page 20606, right column, line 23 - line 26 ---	1-17
Y	J. CELL BIOLOGY vol. 118, no. 3 , August 1992 , ROCKEFELLER UNIV. PRESS, N.Y. , US; pages 541 - 549 P.S. KIM ET AL. 'Transient aggregation of nascent thyroglobulin in the endoplasmic reticulum: relationship to the molecular chaperone, BiP' see page 541, right column, line 13 - line 16 see page 549, right column, line 27 - line 37 --- -/--	1-17

1

C.(Continuation) DOCUMENTS CONSIDE[illegible]TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MOLECULAR BIOLOGY OF THE CELL vol. 3, no. 2 , February 1992 , AMERICAN SOCIETY FOR CELL BIOLOGY; pages 143 - 155 D.T.W. NG ET AL. 'Analysis in vivo of GRP78-BiP/Substrate interactions and their role in induction of the GRP78-BiP gene' see page 152, right column, line 33 - line 35 see page 152, right column, line 42 - line 45	1-17
A	NATURE vol. 337 , 5 January 1989 , MACMILLAN JOURNALS LTD., LONDON, UK; pages 44 - 47 P. GOLOUBINOFF ET AL. 'GroE heat-shock proteins promote assembly of foreign procaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli' see page 45, left column, line 1 - page 47, left column, line 17	1-17
A	TRENDS IN BIOTECHNOLOGY vol. 8, no. 12 , December 1990 , ELSEVIER SCIENCE PUBLISHERS,LTD.,CAMBRIDGE,UK; pages 354 - 358 A.A. GATENBY ET AL. 'Chaperonin assisted polypeptide folding and assembly: implications for the production of functional proteins in bacteria' see page 354, right column, line 1 - line 40	1-17
A	TRENDS IN BIOTECHNOLOGY vol. 8, no. 5 , May 1990 , ELSEVIER SCIENCE PUBLISHERS,LTD.,CAMBRIDGE,UK; pages 126 - 131 A.L. HORWICH ET AL. 'Protein-catalysed protein folding' cited in the application see page 126, left column, line 1 - page 127, left column, line 18; table 1	1-17
P,X	WO,A,93 11248 (CIBA-GEIGY AG) 10 June 1993 see page 7, line 14 - line 21; claims 1-24	1-6,13

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

BNSDOCID: <WO 9408012A1_I >

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 93/09426

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9311248	10-06-93	NL-A- 9102009	16-06-93
		AU-A- 2946192	28-06-93

Form PCT/ISA/210 (patent family annex) (July 1992)